Dynamics of adaptation in an anisotropic phenotype-fitness landscape

F. Hamel [a], F. Lavigne [a,b,c], G. Martin [c] and L. Roques [b]

[a] Aix Marseille Univ, CNRS, Centrale Marseille, I2M, Marseille, France

[b] INRAE, BioSP, 84914, Avignon, France

[c] ISEM (UMR 5554), CNRS, 34095, Montpellier, France

Abstract

We study the dynamics of adaptation of a large asexual population in a n-dimensional phenotypic space, under anisotropic mutation and selection effects. When $n = 1$ or under isotropy assumptions, the 'replicator-mutator' equation is a standard model to describe these dynamics. However, the n-dimensional anisotropic case remained largely unexplored.

We prove here that the equation admits a unique solution, which is interpreted as the phenotype distribution, and we propose a new and general framework to the study of the quantitative behavior of this solution. Our method builds upon a degenerate nonlocal parabolic equation satisfied by the distribution of the 'fitness components', and a nonlocal transport equation satisfied by the cumulant generating function of the joint distribution of these components. This last equation can be solved analytically and we then get a general formula for the trajectory of the mean fitness and all higher cumulants of the fitness distribution, over time. Such mean fitness trajectory is the typical outcome of empirical studies of adaptation by experimental evolution, and can thus be compared to empirical data.

In sharp contrast with the known results based on isotropic models, our results show that the trajectory of mean fitness may exhibit $(n-1)$ plateaus before it converges. It may thus appear 'non-saturating' for a transient but possibly long time, even though a phenotypic optimum exists. To illustrate the empirical relevance of these results, we show that the anisotropic model leads to a very good fit of *Escherichia coli* long-term evolution experiment, one of the most famous experimental dataset in experimental evolution. The two 'evolutionary epochs' that have been observed in this experiment have long puzzled the community: we propose that the pattern may simply stem form a climbing hill process, but in an anisotropic fitness landscape.

1 Introduction

Biological motivation

Understanding the adaptation of asexual organisms (such as viruses, bacteria or cancer cells) under the combined effects of selection and mutation is a fundamental issue in population genetics.

In parallel, the development of experimental evolution (especially in microbes) has made it possible to compare observed dynamics, in the lab, with alternative models or infer various model parameters from the data (for a recent special issue on this subject see [38]). Still, the problem of modelling asexual evolutionary dynamics is inherently complex (discussed in *e.g.* [16]): recurrent mutation changes the key parameters of the dynamical system constantly, competition between numerous, ever changing types must be described, and both mutational and demographic (birth/death) events are stochastic in nature.

Recent models of asexual adaptation seek to follow the dynamics of the full distribution of fitness - the expected reproductive output of a lineage - within populations. Contrarily to other approaches, such as 'origin fixation models' which only follow the expected mean fitness $\overline{m}(t)$ [40], these models do not make a low polymorphism assumption, but in exchange for ignoring or simplifying the stochastic components of the dynamics. The resulting outcome is a partial differential equation (PDE) or an integro-differential equation (IDE), that typically describes the dynamics of the distribution of a single 'trait'. In some cases, this trait may be fitness itself as in [1, 17, 42], leading to equations of the form:

$$\partial_t p(t,m) = \mathcal{M}_1[t,m,p(t,m)] + (m - \overline{m}(t))\, p(t,m),$$

where the variable $m \in \mathbb{R}$ is the fitness and $\mathcal{M}_1$ is a differential or an integral operator describing the effect of mutations on the distribution $p(t,\cdot)$ of fitness. Other models describe the distribution of a given trait $x \in \mathbb{R}$ determining fitness (birth rate, phenotype), as in [2, 8]:

$$\partial_t q(t,x) = \mathcal{M}_2[t,x,q(t,x)] + (m(x) - \overline{m}(t))\, q(t,x),$$

where $m(x)$ is a function which describes the relationship between the trait x and the fitness. Here, $\mathcal{M}_2$ is a differential or an integral operator describing the effect of mutations on the distribution $q(t,\cdot)$ of the trait x. In these two equations, the last term $(m - \overline{m})\, p(t,m)$ (resp. $(m(x) - \overline{m}(t))\, q(t,x)$) corresponds to the effects of selection and will be explained later. Broadly, it emerges when the growth rate $m(x)$ (in the absence of competition) of a genotype with phenotype x depends only on its own phenotype. This corresponds to the classic frequency-independent model of selection.

One established finding from empirical fitness trajectories is that epistasis, namely the fact that the distribution of a mutation's effect on fitness depends on its ancestor's genetic background, must be accounted for to explain the data (*e.g.* [21]). More precisely, fitness trajectories tend to decelerate over time, implying *a priori* that beneficial mutations become weaker and/or rarer as the population adapts (ignoring deleterious ones, which is of course debatable). The question is then: which particular form of epistasis does explain/predict the data best? A common metaphor to explain this decelerating pattern in fitness trajectories is to invoke some form of 'fitness landscape' connecting genotypes or phenotypes with fitness, as in the function $m(x)$ above, with one or several adaptive peak(s) where fitness is maximal (discussed in [11]).

In this view, deceleration in fitness trajectories stems from the hill climbing process of adaptation up a fitness peak. This view is appealing because of its intuitive/visual illustration, but also because of a particular form of landscape, Fisher's geometrical model (FGM; a single peak phenotype-fitness landscape), has shown promising potential when compared to empirical measures of fitness epistasis. In the FGM, a multivariate phenotype at a set of n traits (a vector

$\mathbf{x} \in \mathbb{R}^n$) determines fitness. The most widely used version assumes a quadratic form of the Malthusian fitness function $m(\mathbf{x})$, which decreases away from a single optimum $\mathbf{O} \in \mathbb{R}^n$ [31, 41]. Without loss of generality, the traits can be defined such that the optimum sets the origin of phenotype space:

$$m(\mathbf{x}) = -\frac{\|\mathbf{x}\|^2}{2},$$

with $\|\cdot\|$ the Euclidian norm in $\mathbb{R}^n$. To describe the mutation effects on phenotypes, the standard 'isotropic Gaussian FGM' uses a normal distribution $\mathcal{N}(0, \lambda I_n)$ with $\lambda > 0$ the phenotypic mutational variance at each trait and I_n the identity matrix.

Given the potential complexity of the relationship between genotype, phenotype and fitness, the FGM may appear highly oversimplified. However, several arguments suggest that it can be more than a mere heuristic and may be a relevant, empirically sound null model of adaptation (also discussed in a recent review of the FGM [41]). First [27], the Gaussian FGM can be retrieved from a much less constrained model of phenotype-to-fitness landscape: Random Matrix theory arguments show that it emerges as a limit of highly integrative phenotypic networks, near an optimum (with an arbitrary fitness function and distribution of the original phenotypes). Second, several assumptions of the classic FGM have been tested quite extensively on empirical distributions of mutation effects on fitness, showing its power to fit or predict the observed patterns. For example, the FGM has shown to accurately predict observed distributions of epistasis among random and beneficial mutations in a virus and a bacterium [28], to accurately fit patterns of epistasis among beneficial mutations in a fungus [39] or the pattern of re-adaptation (compensation) from different deleterious mutant backgrounds in a 'recovery' experiment with a bacterium [34]. It also accurately predicted how stressful conditions (lowering the parental fitness) affect the mean and variance of mutation fitness effects [30]. The assumption of a quadratic fitness function is perhaps the most critical to the patterns tested so far: various empirical tests showed that a deviation from the quadratic would actually lead to a lower fit of the data (see [13] and fig. 5 in [30]). Overall, this suggests that, in spite of its apparent simplicity, the FGM with a quadratic fitness function is not so oversimplified: it is a natural limit of a diversity of more complex models and is consistent with various empirical patterns of mutation fitness effects.

On the other hand, this 'fitness peak' view is challenged by the observation [45] that, in the longest evolution experiment ever undertaken (Long Term Evolution Experiment LTEE, in the bacterium *Escherishia coli*), fitness has not reached any maximum after more than 70,000 generations. It has been suggested [19] that this experiment actually shows a 'two epochs' dynamics with an initial (potentially saturating) fitness trajectory and a later and distinct non-saturating dynamics. A similar pattern could be invoked in another mid-term experiment with an RNA virus [33]. Several other experiments did seem to yield a 'true' saturation (plateau) in fitness trajectories over several thousands of generations in *E. coli* (*e.g.* [9, 12, 22]), but they may simply be on a too short timescale to identify subsequent fitness increases. In fact, the LTEE itself did seem to show a strongly saturating pattern after 10,000 generations [24].

Overall, these different insights on epistasis and adaptation trajectories are difficult to reconcile under a single theory: why would a single peak model show a good fit to mutation epistasis over single mutations, or short term fitness trajectories, yet be invalid over the longer term? Is

there really a two 'epochs' dynamics in some long-term trajectories, when and why? A desirable model should reconcile both timescales, describe both epistasis among random single mutations and in long term fitness trajectories. Proposed models that do accurately fit non-saturating trajectories [45] have the drawback that they only focus on beneficial mutations and do not yield any prediction on the distribution of epistasis among random mutations. On the other hand, simple fitness landscapes like the FGM do predict some short term patterns and the full distribution of epistasis among random mutants, but cannot yield a never-ending fitness trajectory (a fitness plateau must be reached at the optimum).

Aim of the paper

In this article, we explore the possibility that an extended version of the FGM be able to fit multiple epochs fitness trajectories. The central extension proposed here is to consider anisotropy in the landscape, in that different underlying traits affect fitness differently and mutate differently. Namely, the fitness function associated with a trait $\mathbf{x} = (x_1, \ldots, x_n)$ is:

$$m(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i \, m_i(\mathbf{x}), \text{ where } m_i(\mathbf{x}) = -\frac{x_i^2}{2}, \tag{1}$$

where the coefficients α_i are positive. The mutation effects on phenotypes are described with an anisotropic Gaussian distribution $\mathcal{N}(0, \Lambda)$, where $\Lambda = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_n)$ is any given positive diagonal matrix.

Using central limit and random matrix theory arguments [27], the FGM can be obtained as a limit of a much less constrained model where high-dimensional phenotypes integrate into a phenotypic network to a smaller set of traits that directly determine fitness, with an optimum. The resulting FGM, however, is not necessarily isotropic, it can also show a single dominant direction in phenotype space (that affects fitness much more) with all other directions remaining approximately equivalent. Our initial intuition is that, in these conditions, adaptation along the dominant direction will drive the early dynamics while a second 'epoch' will be visible when adaptation only proceeds along the remaining dimensions. Therefore, it seems natural to explore how such a model would fare when compared to the fitness trajectories of the LTEE. Yet, this requires deriving the fitness dynamics resulting from such a complex mutational model, in the presence of mutation and competition between asexual lineages. This is the aim of the present paper.

The existence of a phenotype optimum has been taken into account in recent PDE and IDE approaches (second class of models alluded to above). For instance, in [2, 3], the fitness depends on a single phenotypic trait $x \in \mathbb{R}$, through a function $m(x)$ which admits an optimum. Extending these works to take into account the dependence of the fitness on several traits appears as a natural question; especially since we know that the number of traits affected by mutation and subject to selection critically affects the evolutionary dynamics [32, 44]. So far, and to the best of our knowledge, such mathematical models that take into account $n-$dimensional phenotypes together with the existence of a phenotype optimum always assume an isotropic dependence between the traits $\mathbf{x} \in \mathbb{R}^n$ and the fitness, and an isotropic effect of mutations on phenotypes (see [18] for an IDE approach and [32] for an approach based on the analysis of a

PDE satisfied by a moment generating function of the fitness distribution). Our goal here is to extend these approaches in order to characterize the dynamics of adaptation in a n-dimensional phenotypic space, without making an isotropy assumption. For the sake of simplicity here, we ignore any stochastic component of the dynamics, but note that the same equations can be obtained explicitly as a deterministic limit of a stochastic model of mutation and growth [32].

Model assumptions and definitions

We assume that a phenotype is a set of n biological traits, which is represented by a vector $\mathbf{x} \in \mathbb{R}^n$. To describe the evolution of the phenotypic distribution we begin with an intuitive description of the concept of fitness. In an asexual population of K types of phenotypes $\mathbf{x}_1, \ldots, \mathbf{x}_K$, with $r_{\max} > 0$ the growth rate of the optimal phenotype, the *Malthusian fitness, relative to the optimal phenotype* is denoted $m_j = m(\mathbf{x}_j) \leq 0$ for a phenotype $\mathbf{x}_j \in \mathbb{R}^n$. It is defined by the following relation: $N_j'(t) = (r_{\max} + m_j) N_j(t)$, with $N_j(t)$ the abundance of the phenotype $\mathbf{x}_j$ at time t. When we sum these equations over all indexes $j = 1, \ldots, K$, we obtain an ordinary differential equation for the total population size $N(t) = \sum_{j=1}^{K} N_j(t)$ at time t: $N'(t) = (r_{\max} + \overline{m}(t)) N(t)$, where the quantity $\overline{m}(t) = \sum_{j=1}^{K} m_j N_j(t)/N(t)$ is the mean relative fitness in the population at time t. Now if we turn to the distribution of the phenotype frequencies $q(t, \mathbf{x}_j) = N_j(t)/N(t)$, we get the partial differential equation:

$$\partial_t q(t, \mathbf{x}_j) = (m(\mathbf{x}_j) - \overline{m}(t))\, q(t, \mathbf{x}_j). \tag{2}$$

This equation can be generalized to a continuous distribution of phenotype frequencies (see *e.g.* [42]), as we assume in the sequel, with:

$$\overline{m}(t) = \int_{\mathbb{R}^n} m(\mathbf{x})\, q(t, \mathbf{x})\, d\mathbf{x}. \tag{3}$$

If $m(\mathbf{x}) < \overline{m}(t)$, then (2) implies that the frequency of the phenotype $\mathbf{x}$ decreases, whereas if $m(\mathbf{x}) > \overline{m}(t)$, then the frequency increases.

We recall that mutation's effects on phenotypes are assumed to follow an anisotropic Gaussian distribution $\mathcal{N}(0, \Lambda)$, with $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ a positive diagonal matrix. Assuming a small mutational variance $\max \lambda_i << 1$, and a mutation rate U, these mutational effects can be approximated by an elliptic operator $\sum_{i=1}^{n} (\mu_i^2/2) \partial_{ii}$, where $\mu_i = \sqrt{U \lambda_i} > 0$ and ∂_{ii} denotes the second order partial derivative with respect to the i-th coordinate of $\mathbf{x}$ (or later $\mathbf{m}$ as in (17) below). We refer to Appendix A for further details on the derivation of this diffusion approximation. Intuitively, the regime where it applies is the 'Weak Selection Strong Mutation' (WSSM) regime (high rate of small effect mutations) where a wide diversity of lineages accumulate mutations and co-segregate at all times. The extreme alternative is the Strong Selection Weak Mutation (SSWM) regime where few mutations of large effect invade successively and the population is effectively close to monomorphic at all times. This latter regime will not be treated here.

Overall, the corresponding PDE describing the dynamics of the phenotype distribution $q(t, \mathbf{x})$, under the combined effects of selection and mutation, is:

$$\partial_t q(t, \mathbf{x}) = \sum_{i=1}^{n} \frac{\mu_i^2}{2} \partial_{ii} q(t, \mathbf{x}) + (m(\mathbf{x}) - \overline{m}(t))\, q(t, \mathbf{x}), \ t > 0, \ \mathbf{x} \in \mathbb{R}^n, \tag{4}$$

with $\overline{m}(t)$ defined by (3). Recent studies [2] have already treated the case $n = 1$ (see also [3] for more general fitness functions). In this paper, we consider the general case $n \geq 1$. Without loss of generality, we may assume in the sequel that $\alpha_1 = \ldots = \alpha_n = 1$ in (5), up to a scaling of the phenotype space ($\tilde{q}(t, \mathbf{x}) = (\prod_{i=1}^n \alpha_i^{1/2})\ q(t, x_1/\sqrt{\alpha_1}, \ldots, x_n/\sqrt{\alpha_n})$). This leads to:

$$m(\mathbf{x}) = -\frac{\|\mathbf{x}\|^2}{2}, \tag{5}$$

with $\|\cdot\|$ the standard Euclidian norm in $\mathbb{R}^n$. Similarly, we could remove the anisotropy in the mutation effects, up to an other scaling of the phenotype space (defined this time by $\tilde{q}(t, \mathbf{x}) = (\prod_{i=1}^n \mu_i^{-1}) q(t, \mu_1 x_1, \ldots, \mu_n x_n)$), but in this case the coefficients α_i would be replaced by $\alpha_i \sqrt{\mu_i}$ and hence cannot be taken all equal to 1. Note that, for the sake of simplicity, we assumed a diagonal corvariance matrix Λ and that $m(\mathbf{x})$ was a linear combination of the fitness components $m_i(\mathbf{x}) = -x_i^2/2$. More general covariance matrices Λ and quadratic forms $m(\mathbf{x})$ could have been considered as well using the transformations of the phenotype space presented in [29].

In the one-dimensional case, an equation of the form (4) with a general nonlocal reaction term of the form $m(x) - \int_{\mathbb{R}} m(z) q(t, z) dz$ has been studied in [3]. Under the assumption that $m(x)$ tends to $-\infty$ as x tends to $\pm\infty$, the authors have established a formula for q involving the eigenelements of the operator $\mathcal{H}_1 = \frac{\mu^2}{2} d^2/dx^2 + m(x)$. The formula can be made more explicit for our choice of m, $m(x) = -x^2/2$ [2]. However, the method used in [2], which consists in reducing the equation (4) to the heat equation thanks to changes of variables based on Avron-Herbst formula and generalized Lens transform, cannot be directly applied in our $n-$dimensional anisotropic framework. Recently, another method based on constrained Hamilton-Jacobi equations has been developed to study the evolution of phenotypically structured populations, with integral or differential mutation operators (*e.g.*, [6, 10, 15, 25, 35]). This method assumes a small mutation parameter $\mu^2/2$ (the terms $\mu_i^2/2$ in (4)) of order $\varepsilon \ll 1$, and is based on a scaling $t \to t/\varepsilon$. Thus, it typically describes asymptotic evolutionary dynamics, at large times and in a 'small mutation' regime. Note however that, as $\mu^2 = U\lambda$, with U the mutation rate and λ the mutational variance (at each trait), it encompasses the cases where the mutation rate is not small, provided that $\lambda \ll 1/U$. This method should apply with anisotropic mutation operators, as in (4). However, to the best of our knowledge, it cannot lead to explicit transient trajectories of adaptation, which is the real objective of our work. We use here a completely different approach.

In order to understand the dynamics of adaptation, an important quantity is of course the *fitness distribution* $p(t, m)$, such that $p(t, m)\, dm$ is the pushforward measure of the measure $q(t, \mathbf{x})\, d\mathbf{x}$ by the map $\mathbf{x} \mapsto -\|\mathbf{x}\|^2/2$, and the *mean fitness* $\overline{m}(t)$. In the isotropic case [18], the authors directly focused on the equation satisfied by $p(t, m)$ and not by $q(t, \mathbf{x})$. Whereas (4) is quadratic into the variables $x_1, \ldots, x_n$, the equation satisfied by $p(t, m)$ is linear with respect to m, which makes possible the derivation of some PDEs satisfied by generating functions of $p(t, m)$. Here, due to the anisotropy, the dynamics of $p(t, m)$ is not summarized by a single one-dimensional PDE. Instead, we define the fitness components $\mathbf{m} = (m_1, \ldots, m_n)$ and the *joint distribution of the fitness components* $\mathbf{p}(t, \mathbf{m})$ such that $\mathbf{p}(t, \mathbf{m})\, d\mathbf{m}$ is the pushforward measure of the measure $q(t, \mathbf{x})\, d\mathbf{x}$ by the map $\mathbf{x} \mapsto (-x_1^2/2, \cdots, -x_n^2/2)$ (see Proposition 2.3 for more details). As fitness is the sum of the fitness components, the distributions $\mathbf{p}$ and p are

linked by:

$$p(t,m) = \int_{\mathbb{R}_-^{n-1}\cap\{m\le\sum_{i=1}^{n-1} m_i\}} \mathbf{p}\left(t, m_1, \dots, m_{n-1}, m - \sum_{i=1}^{n-1} m_i\right) dm_1 \dots dm_{n-1}, \tag{6}$$

where $\mathbb{R}_- = (-\infty, 0]$.

Thus the mean fitness $\overline{m}(t)$ can be easily connected with these distributions:

$$\overline{m}(t) = \int_{\mathbb{R}^n} m(\mathbf{x}) q(t,\mathbf{x}) d\mathbf{x} = \sum_{i=1}^{n} \int_{\mathbb{R}_-^n} m_i \mathbf{p}(t,\mathbf{m}) d\mathbf{m} = \int_{\mathbb{R}_-} m p(t,m) dm,$$

see also (15) below.

This paper is organized as follows. The results are presented in Section 2, and their proofs are presented in Section 4. More precisely, Section 2.1 is dedicated to the analysis of the time-dependent problem (4). We begin with the existence and uniqueness of the solution of the Cauchy problem. We then give an explicit formula for $q(t,\mathbf{x})$ in the particular case of a Gaussian initial distribution of phenotypes $q_0(\mathbf{x})$. Then, we derive a nonlocal degenerate parabolic equation satisfied by $\mathbf{p}(t,\mathbf{m})$, and the equation satisfied by its cumulant generating function (CGF). Solving the equation satisfied by the CGF, we derive an explicit formula for $\overline{m}(t)$. Then, in Section 2.2, we study the long time behavior and the stationary states of (4). We shall see that the distribution of the fitness components $\mathbf{p}(t,\mathbf{m})$ converges towards a distribution $\mathbf{p}_\infty(\mathbf{m})$ as $t \to +\infty$, and we give an explicit formula for $\mathbf{p}_\infty(\mathbf{m})$. Lastly, in Section 2.3 we show that including anisotropy in the models may help to understand experimental trajectories of fitnesses, such as those obtained in the famous experiment of Richard Lenski [23, 24, 45].

2 Main results

2.1 The time-dependent problem

Solution of the Cauchy problem associated with equation (4) **for** $q(t,\mathbf{x})$

We first show that the Cauchy problem admits a unique solution. We need the following assumption on the initial distribution q_0:

$$q_0 \in C^{2+\alpha}(\mathbb{R}^n), \tag{7}$$

for some $\alpha \in (0,1)$, that is, $\|q_0\|_{C^{2+\alpha}(\mathbb{R}^n)} < +\infty$. Additionally, we assume that:

$$\int_{\mathbb{R}^n} q_0(\mathbf{x}) d\mathbf{x} = 1, \tag{8}$$

and there exists a function $g : \mathbb{R}_+ \to \mathbb{R}_+$ (with $\mathbb{R}_+ = [0,+\infty)$) such that:

$$g \text{ is non-increasing}, \quad 0 \le q_0 \le g(\|\cdot\|) \text{ in } \mathbb{R}^n, \ \mathbf{x} \mapsto m(\mathbf{x}) g(\|\mathbf{x}\|) \text{ is bounded in } \mathbb{R}^n, \quad \text{and } \int_{\mathbb{R}^n} |m(\mathbf{x})|\, g(\|\mathbf{x}\|)\, d\mathbf{x} < +\infty. \tag{9}$$

These assumptions are made throughout the paper, and are therefore not repeated in the statements of the results below.

We can now state an existence and uniqueness result for the distribution of phenotypes.

Theorem 2.1. *There exists a unique nonnegative solution $q \in C^{1,2}(\mathbb{R}_+ \times \mathbb{R}^n)$ of:*

$$\left\{ \begin{array}{rcl} \partial_t q(t,\mathbf{x}) & = & \displaystyle\sum_{i=1}^n \frac{\mu_i^2}{2} \partial_{ii} q(t,\mathbf{x}) + (m(\mathbf{x}) - \overline{m}(t))\, q(t,\mathbf{x}), \quad t \geq 0,\ \mathbf{x} \in \mathbb{R}^n, \\ q(0,\mathbf{x}) & = & q_0(\mathbf{x}),\ \mathbf{x} \in \mathbb{R}^n, \end{array} \right. \tag{10}$$

such that $q \in L^\infty((0,T) \times \mathbb{R}^n)$ for all $T > 0$, and the function:

$$t \mapsto \overline{m}(t) = \int_{\mathbb{R}^n} m(\mathbf{x})\, q(t,\mathbf{x})\, d\mathbf{x},$$

is real-valued and continuous in $\mathbb{R}_+$. Moreover, we have:

$$\forall\, t \geq 0, \quad \int_{\mathbb{R}^n} q(t,\mathbf{x})\, d\mathbf{x} = 1.$$

The next result gives an explicit solution of (10) in the particular case where the phenotypes are initially Gaussian-distributed.

Corollary 2.2. Assume that the initial distribution of phenotype frequencies is Gaussian, that is,

$$\forall\, \mathbf{x} \in \mathbb{R}^n,\ q_0(\mathbf{x}) = (2\pi)^{-n/2} \Big(\prod_{i=1}^n (s_i^0)^{-1/2} \Big) \exp\left(- \sum_{i=1}^n \frac{(x_i - \overline{q}_i^0)^2}{2 s_i^0} \right), \tag{11}$$

for some parameters $\overline{q}_i^0 \in \mathbb{R}$ and $s_i^0 > 0$. Then the solution $q(t,\mathbf{x})$ of the Cauchy problem (10) is Gaussian at all time:

$$\forall\, t \geq 0,\ \forall\, \mathbf{x} \in \mathbb{R}^n,\ q(t,\mathbf{x}) = (2\pi)^{-n/2} \Big(\prod_{i=1}^n (s_i(t))^{-1/2} \Big) \exp\left(- \sum_{i=1}^n \frac{(x_i - \overline{q}_i(t))^2}{2 s_i(t)} \right), \tag{12}$$

with:

$$\overline{q}_i(t) = \frac{\mu_i \overline{q}_i^0}{\mu_i \cosh(\mu_i t) + s_i^0 \sinh(\mu_i t)}, \qquad \text{and} \qquad s_i(t) = \mu_i \frac{\mu_i \sinh(\mu_i t) + s_i^0 \cosh(\mu_i t)}{\mu_i \cosh(\mu_i t) + s_i^0 \sinh(\mu_i t)}. \tag{13}$$

Moreover, we have:

$$\overline{m}(t) = - \sum_{i=1}^n \frac{\overline{q}_i^2(t) + s_i(t)}{2}.$$

We also note, in Corollary 2.2, that the distribution $q(t,\mathbf{x})$ converges to a Gaussian distribution with mean $\overline{q}_\infty = 0$ and variances $s_{\infty,i} = \mu_i$, as $t \to +\infty$.

The determination of an explicit formula for $q(t,\mathbf{x})$ becomes more involved when the initial distribution q_0 is not a Gaussian. In this case, we study the equation satisfied by the distribution of the fitness components $\mathbf{p}(t,\mathbf{m})$ (see Introduction) and, as a by-product, we derive an explicit formula for $\overline{m}(t)$.

A degenerate parabolic PDE satisfied by $\mathbf{p}(t,\mathbf{m})$

Our objective here is to derive an equation for $\mathbf{p}(t,\mathbf{m})$ that only involves linear dependencies with respect to the coefficients m_i, and holds for general initial phenotype distributions q_0, which may not be Gaussian.

First, we express the distribution of the fitness components $\mathbf{p}(t,\mathbf{m})$ in terms of the distribution of phenotypes $q(t,\mathbf{x})$ given in Theorem 2.1.

Proposition 2.3. *For all $t \geq 0$ and $\mathbf{m} = (m_1, \dots, m_n) \in (\mathbb{R}_-^*)^n$, there holds:*

$$\mathbf{p}(t, \mathbf{m}) = \frac{2^{-n/2}}{\sqrt{|m_1 \cdots m_n|}} \sum_{\varepsilon = (\varepsilon_1, \dots, \varepsilon_n) \in \{\pm 1\}^n} q(t, \mathbf{x}^\varepsilon(\mathbf{m})), \tag{14}$$

with $\mathbf{x}^\varepsilon(\mathbf{m}) = (\varepsilon_1 \sqrt{-2m_1}, \dots, \varepsilon_n \sqrt{-2m_n}) \in \mathbb{R}^n$. Furthermore, we have:

$$\forall\, t \geq 0, \quad \int_{\mathbb{R}_-^n} \mathbf{p}(t, \mathbf{m})\, d\mathbf{m} = 1, \qquad \text{and} \qquad \overline{m}(t) = \sum_{i=1}^n \int_{\mathbb{R}_-^n} m_i \mathbf{p}(t, \mathbf{m})\, d\mathbf{m}, \tag{15}$$

where all above integrals are convergent.

It also turns out that the expression (14) becomes simpler when q satisfies some symmetry properties. In the sequel, for a given function $f \in C(\mathbb{R}^n)$, we define its $\#$−symmetrization $f^\# \in C(\mathbb{R}^n)$ by:

$$\forall\, \mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n, \; f^\#(\mathbf{x}) = 2^{-n} \sum_{\varepsilon = (\varepsilon_1, \dots, \varepsilon_n) \in \{\pm 1\}^n} f(\varepsilon_1 x_1, \dots, \varepsilon_n x_n).$$

From the symmetries inherent to (4), it is easy to check that, if q is the solution of (4) with initial condition q_0 satisfying the conditions of Theorem 2.1, then $q^\#$ is the solution of (4) with initial condition $q_0^\#$, and $q(t, \cdot)$ and $q^\#(t, \cdot)$ have the same mean fitness $\overline{m}(t)$ at every time $t \geq 0$. Furthermore, using the expression (14), we observe that $\mathbf{p}(t, \mathbf{m})$ can be described in terms of the $\#$−symmetrization of $q(t, \cdot)$:

$$\mathbf{p}(t, \mathbf{m}) = \frac{2^{n/2}}{\sqrt{|m_1 \cdots m_n|}} q^\#(t, \mathbf{x}^{\mathbf{1}}(\mathbf{m})), \tag{16}$$

for every $t \geq 0$ and $\mathbf{m} \in (\mathbb{R}_-^*)^n$, with:

$$\mathbf{x}^{\mathbf{1}}(\mathbf{m}) = (\sqrt{-2m_1}, \dots, \sqrt{-2m_n}) \in \mathbb{R}_+^n.$$

This function $\mathbf{p}(t, \mathbf{m})$ satisfies a degenerate parabolic equation:

Theorem 2.4. *The distribution function of the fitness components $\mathbf{p}$ is a classical $C^{1,2}(\mathbb{R}_+ \times (\mathbb{R}_-^*)^n)$ solution of:*

$$\partial_t \mathbf{p}(t, \mathbf{m}) = \sum_{i=1}^n \mu_i^2 |m_i|\, \partial_{ii} \mathbf{p}(t, \mathbf{m}) - \frac{3}{2} \sum_{i=1}^n \mu_i^2\, \partial_i \mathbf{p}(t, \mathbf{m}) + \left(\sum_{i=1}^n m_i - \overline{m}(t) \right) \mathbf{p}(t, \mathbf{m}), \tag{17}$$

for $t \geq 0$ and $\mathbf{m} \in (\mathbb{R}_-^*)^n$, with initial condition:

$$\mathbf{p}_0(\mathbf{m}) = \frac{2^{n/2}}{\sqrt{|m_1 \cdots m_n|}} q_0^\#(\mathbf{x}^{\mathbf{1}}(\mathbf{m})). \tag{18}$$

The equation (17) is degenerate in the sense that the operator in the right-hand side is not uniformly elliptic, as at least one of the coefficients $\mu_i^2 |m_i|$ in front of the second order differential terms vanishes at the boundary of $(\mathbb{R}_-^*)^n$.

In the isotropic case ($\mu_i = \mu > 0$ for all i), it is also possible to derive a scalar equation for the distribution of fitness $p(t, m)$, defined by (6).

Theorem 2.5. (*Isotropic case*) *If $\mu_i = \mu$ for all $1 \le i \le n$, then the fitness distribution $p(t,m)$ is a classical $C^{1,2}(\mathbb{R}_+ \times \mathbb{R}_-^*)$ solution of:*

$$\partial_t p(t,m) = -\mu^2\, m\, \partial_{mm} p(t,m) + \mu^2 \left(\frac{n}{2} - 2\right) \partial_m p(t,m) + (m - \overline{m}(t))\, p(t,m), \tag{19}$$

for $t \ge 0$ and $m < 0$, with initial condition,

$$p_0(m) = \int_{\mathbb{R}_-^{n-1} \cap \{m \le \sum_{i=1}^{n-1} m_i\}} \mathbf{p}_0 \left(m_1, \dots, m_{n-1}, m - \sum_{i=1}^{n-1} m_i\right) dm_1 \dots dm_{n-1}. \tag{20}$$

As expected, the equations (17) and (19) only involve linear combinations of the coefficients m_i. This allows us to derive simpler equations satisfied by the generating functions of $\mathbf{p}(t,\mathbf{m})$ and $p(t,m)$.

Generating functions

We define the *moment generating functions* (MGFs) $M_{\mathbf{p}}$ and M_p of $\mathbf{p}$ and p and their logarithms – the *cumulant generating functions* (CGFs) – $C_{\mathbf{p}}$ and C_p by:

$$M_{\mathbf{p}}(t,\mathbf{z}) = \int_{\mathbb{R}_-^n} e^{\mathbf{z}\cdot\mathbf{m}} \mathbf{p}(t,\mathbf{m})\, d\mathbf{m}, \quad M_p(t,z) = \int_{-\infty}^0 e^{zm} p(t,m)\, dm, \tag{21}$$

and:

$$C_{\mathbf{p}}(t,\mathbf{z}) = \log M_{\mathbf{p}}(t,\mathbf{z}), \quad C_p(t,z) = \log M_p(t,z), \tag{22}$$

for $t \ge 0$, $\mathbf{z} \in \mathbb{R}_+^n$ and $z \ge 0$. The integrals are well defined because the components m_i are all negative. Furthermore, it follows from (15) and the nonnegativity of $\mathbf{p}$ that, for each $t \ge 0$, the functions $M_{\mathbf{p}}(t,\cdot)$ and $C_{\mathbf{p}}(t,\cdot)$ are of class $C^\infty((\mathbb{R}_+^*)^n) \cap C^1(\mathbb{R}_+^n)$, while the functions $M_p(t,\cdot)$ and $C_p(t,\cdot)$ are of class $C^\infty(\mathbb{R}_+^*) \cap C^1(\mathbb{R}_+)$.

The following result gives the equation satisfied by $C_{\mathbf{p}}$.

Theorem 2.6. *The cumulant generating function $C_{\mathbf{p}}$ of $\mathbf{p}$ is of class $C^{1,1}(\mathbb{R}_+ \times \mathbb{R}_+^n)$* [1] *and it solves:*

$$\begin{cases} \partial_t C_{\mathbf{p}}(t,\mathbf{z}) &= A(\mathbf{z}) \cdot \nabla C_{\mathbf{p}}(t,\mathbf{z}) - b(\mathbf{z}) - \overline{m}(t),\ t \ge 0,\ \mathbf{z} \in \mathbb{R}_+^n, \\ C_{\mathbf{p}}(0,\mathbf{z}) &= C_{\mathbf{p}_0}(\mathbf{z}),\ \mathbf{z} \in \mathbb{R}_+^n, \end{cases} \tag{23}$$

where:

$$A(\mathbf{z}) = (1 - \mu_1^2 z_1^2, \dots, 1 - \mu_n^2 z_n^2) \in \mathbb{R}^n, \quad b(\mathbf{z}) = \sum_{i=1}^n \frac{\mu_i^2}{2} z_i \in \mathbb{R}, \quad \overline{m}(t) = \sum_{i=1}^n \partial_i C_{\mathbf{p}}(t,\mathbf{O}), \tag{24}$$

and $\nabla C_{\mathbf{p}}(t,\mathbf{z})$ denotes the gradient of $C_{\mathbf{p}}$ with respect to the variable $\mathbf{z}$.

As a corollary of this theorem, we derive an equation satisfied by C_p in the isotropic case.

Corollary 2.7. (Isotropic Case) If $\mu_i = \mu$ for all $1 \le i \le n$, then the cumulant generating function C_p of the fitness distribution p is a $C^{1,1}(\mathbb{R}_+ \times \mathbb{R}_+)$ solution of:

$$\begin{cases} \partial_t C_p(t,z) &= (1 - \mu^2 z^2) \partial_z C_p(t,z) - \frac{n}{2} \mu^2 z - \overline{m}(t),\ t \ge 0,\ z \in \mathbb{R}_+, \\ C_p(0,z) &= C_{p_0}(z),\ z \in \mathbb{R}_+, \end{cases} \tag{25}$$

where p_0 is the initial fitness distribution given in (20), and $\overline{m}(t) = \partial_z C_p(t,0)$.

1. This means that the partial derivatives of $C_{\mathbf{p}}$ with respect to the variables t and $\mathbf{z}$ exist and are continuous in $\mathbb{R}_+ \times \mathbb{R}_+^n$.

Note that the equation (25) is directly obtained as a by-product of (23), without using the equation (19) satisfied by $p(t,m)$. Each of these two equations (23) and (25) has a unique solution which can be computed explicitly, leading to a formula for $\overline{m}(t)$. We refer to Section 2.3 for an application of this result.

In the Introduction, we mentioned that the coefficients μ_i^2 can be interpreted as the product between the mutation rate U and the variance λ_i at the i-th trait. In the isotropic case, $\mu_i^2 = \mu^2 = U\,\lambda$. Thus we have retrieved the equation mentioned in [32, Eq. (E5) in Appendix E].

The last two results of this section provide some explicit expressions of $C_{\mathbf{p}}(t,\mathbf{z})$ and $C_p(t,z)$ when $\mathbf{z}$ and z are close enough to $\mathbf{O}$ and 0.

Proposition 2.8. *The cumulant generating function $C_{\mathbf{p}}$ of $\mathbf{p}$ is given by, for all $t \geq 0$ and $\mathbf{z} = (z_1,\dots,z_n) \in [0,1/\mu_1) \times \cdots \times [0,1/\mu_n)$,*

$$C_{\mathbf{p}}(t,\mathbf{z}) = \frac{1}{2}\sum_{i=1}^{n} \log\left[\frac{\cosh(\mu_i t)\,\cosh(\operatorname{atanh}(\mu_i z_i))}{\cosh(\mu_i t + \operatorname{atanh}(\mu_i z_i))}\right] + C_{\mathbf{p}_0}(\psi(t,\mathbf{z})) - C_{\mathbf{p}_0}(\psi(t,\mathbf{O})), \tag{26}$$

with:

$$\psi(t,\mathbf{z}) = (\psi_1(t,\mathbf{z}),\dots,\psi_n(t,\mathbf{z})), \qquad \text{and} \qquad \psi_j(t,\mathbf{z}) = \frac{1}{\mu_j}\tanh\left(\mu_j t + \operatorname{atanh}(\mu_j z_j)\right). \tag{27}$$

Moreover, for all $t \geq 0$, we have:

$$\overline{m}(t) = \sum_{i=1}^{n}\Big[(1-\tanh^2(\mu_i t))\;\partial_i C_{\mathbf{p}_0}(\psi(t,\mathbf{O})) - \frac{\mu_i}{2}\tanh(\mu_i t)\Big]. \tag{28}$$

Corollary 2.9. (Isotropic case) If $\mu_i = \mu$ for all $1 \leq i \leq n$, then the cumulant generating function C_p of p is given by:

$$C_p(t,z) = \frac{n}{2}\log\left(\frac{\cosh(\mu\,t)\cosh(\operatorname{atanh}(\mu\,z))}{\cosh(\mu\,t + \operatorname{atanh}(\mu\,z))}\right) + C_{p_0}(\varphi(t,z)) - C_{p_0}(\varphi(t,0)), \tag{29}$$

for $t \geq 0$ and $0 \leq z < 1/\mu$, with $\varphi(t,z) = (1/\mu)\tanh\big(\mu\,t + \operatorname{atanh}(\mu\,z)\big)$. Moreover, we have:

$$\overline{m}(t) = (1-\tanh^2(\mu t))\,\partial_z C_{p_0}\Big(\frac{\tanh(\mu t)}{\mu}\Big) - \frac{n\mu}{2}\tanh(\mu t). \tag{30}$$

2.2 Long time behavior and stationary states

We are here interested in the long time behavior of the solutions of (4) and (17). We begin with a result on the convergence of the solution of (17) at $t \to +\infty$.

Theorem 2.10. *Let $\mathbf{p}$ and $\overline{m}(t)$ be as in the previous section. Then:*

(i) $\mathbf{p}(t,\cdot)$ weakly converges in $(\mathbb{R}_-^*)^n$ to $\mathbf{p}_\infty$ as $t \to +\infty$, where:

$$\mathbf{p}_\infty(\mathbf{m}) = \frac{1}{\pi^{n/2}\sqrt{\mu_1\cdots\mu_n}\sqrt{|m_1\cdots m_n|}}\exp\left(\sum_{i=1}^{n}\frac{m_i}{\mu_i}\right) \quad \text{for all } m \in (\mathbb{R}_-^*)^n, \tag{31}$$

in the sense that $\int_{(\mathbb{R}_-^*)^n}\mathbf{p}(t,\mathbf{m})\,\phi(\mathbf{m})\,d\mathbf{m} \to \int_{(\mathbb{R}_-^*)^n}\mathbf{p}_\infty(\mathbf{m})\,\phi(\mathbf{m})\,d\mathbf{m}$ as $t \to +\infty$ for every test function $\phi \in C_c^\infty((\mathbb{R}_-^*)^n)$;

(ii) $\overline{m}(t) \to \overline{m}_\infty = -\sum_{i=1}^{n} \frac{\mu_i}{2}$ *as* $t \to +\infty$ *and* $\overline{m}_\infty = \sum_{i=1}^{n} \int_{\mathbb{R}_-^n} m_i \mathbf{p}_\infty(\mathbf{m}) d\mathbf{m}$*;*

(iii) the function $\mathbf{p}_\infty$ *is a classical* $C^2((\mathbb{R}_-^*)^n)$ *solution of:*

$$0 = \sum_{i=1}^{n} \mu_i^2 |m_i| \partial_{ii} \mathbf{p}_\infty(\mathbf{m}) - \frac{3}{2} \sum_{i=1}^{n} \mu_i^2 \partial_i \mathbf{p}_\infty(\mathbf{m}) + \left(\sum_{i=1}^{n} m_i - \overline{m}_\infty \right) \mathbf{p}_\infty(\mathbf{m}). \tag{32}$$

In the isotropic case, we retrieve the result of [32] in the WSSM case (Weak Selection and Strong Mutation), which says that the fitnesses are asymptotically distributed according to the symetrized Gamma distribution $-\Gamma(n/2, \mu)$, with $\mu = \sqrt{U\lambda}$:

Corollary 2.11. (Isotropic case) If $\mu_i = \mu$ for all $1 \le i \le n$, then $p(t, \cdot)$ weakly converges in $\mathbb{R}_-^*$ to p_∞ as $t \to +\infty$, where:

$$p_\infty(m) = \frac{|m|^{\frac{n}{2}-1}}{\Gamma(n/2)\, \mu^{n/2}} \exp\left(\frac{m}{\mu}\right) \quad \text{for all } m < 0,$$

and $\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt$ is the standard Gamma function.

Thanks to the previous two results, we get the asymptotic behavior of the phenotype distribution in the symmetric case.

Corollary 2.12. If q_0 is #-symmetric in the sense that $q_0 = q_0^{\#}$, then $q(t, \cdot) = q^{\#}(t, \cdot)$ weakly converges to q_∞ as $t \to +\infty$ on $\mathbb{R}^n$ where, for all $x \in \mathbb{R}^n$,

$$q_\infty(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} \sqrt{\mu_1 \cdots \mu_n}} \exp\left(- \sum_{i=1}^{n} \frac{x_i^2}{2\mu_i} \right), \tag{33}$$

and $\overline{m}(t) \to \overline{m}_\infty = \int_{\mathbb{R}^n} m(\mathbf{x})\, q_\infty(\mathbf{x})\, d\mathbf{x}$ as $t \to +\infty$.

We note that q_∞ is a classical positive stationary state of (4), *i.e.*, it satisfies the following equation:

$$\sum_{i=1}^{n} \frac{\mu_i^2}{2} \partial_{ii} q_\infty(\mathbf{x}) + m(\mathbf{x})\, q_\infty(\mathbf{x}) = \overline{m}_\infty\, q_\infty(\mathbf{x}) \quad \text{for all } \mathbf{x} \in \mathbb{R}^n.$$

Thus, the distribution $q(t, \mathbf{x})$ and the mean fitness $\overline{m}(t)$ converge to the principal eigenfunction (resp. eigenvalue) of the operator $\mathcal{H}_n = \sum_{i=1}^{n} \frac{\mu_i^2}{2} \partial_{ii} + m(\mathbf{x})$.

In the 1D case ($x \in \mathbb{R}$), the results in [3] imply that $q(t, x)$ can be written in terms of the eigenelements (λ_k, ϕ_k) of the operator $\mathcal{H}_1 = -\frac{\mu^2}{2} d^2/dx^2 - m(x)$. Namely,

$$q(t, x) = K(t) \sum_{k=0}^{\infty} (q_0, \phi_k)_{L^2(\mathbb{R})}\, \phi_k(x)\, e^{-\lambda_k t},$$

where $K(t)$ is such that $q(t, \cdot)$ sums to 1, and $(\cdot, \cdot)_{L^2(\mathbb{R})}$ is the standard scalar product in $L^2(\mathbb{R})$. As a corollary, the authors of [3] obtained the convergence of the distribution $q(t, x)$ and of the mean fitness $\overline{m}(t)$ to the principal eigenfunction (resp. eigenvalue) of $\mathcal{H}_1$, namely ϕ_0 (resp. λ_0). Thus, in the 1D case, their result is stronger than that of Corollary 2.12, as the convergence of $q(t, x)$ occurs in $L^p(\mathbb{R}^n)$, for all $1 \le p \le +\infty$, and does not require q_0 to be #-symmetric. Based on the spectral properties of the operator $\mathcal{H}_n$ [20], we conjecture that their results can be extended to the anisotropic multidimensional framework considered here, meaning that the convergence result in Corollary 2.12 remains true when the initial condition q_0 is not #-symmetric.

2.3 Effect of anisotropy: numerical computations and connection with *Escherichia coli* long-term evolution experiment

The objective of this section is to illustrate the importance of taking anisotropy into account when modelling adaptation trajectories. Isotropic models [32] lead to regularly saturating trajectories of $\overline{m}(t)$ with a plateau, *i.e.* a single 'epoch'. Here, we show that, in the presence of anisotropy, the trajectory of $\overline{m}(t)$ can exhibit several plateaus before reaching a stable level close to $\overline{m}_\infty$. Thus, the dynamics of adaptation can show several evolutionary 'epochs', as those observed in the *E. coli* long-term evolution experiment [19], corresponding to different time-scales at which adaptation occurs.

For the sake of simplicity of the computations, and although the existence and uniqueness results of Section 2.1 were only obtained with continuous initial distributions of phenotypes, we assume a Dirac initial distribution of the phenotypes. Namely, we assume that $q_0 = \delta_{\mathbf{x}_0}$ with $\mathbf{x}_0 = (x_{0,1}, \dots, x_{0,n}) \in \mathbb{R}^n$. This corresponds to an initially clonal population. This hypothesis implies that the initial distribution of the fitness components and the initial CGF are respectively given by:

$$\mathbf{p}_0 = \delta_{\mathbf{m}(\mathbf{x}_0)}, \quad \text{and} \quad C_{\mathbf{p}_0}(\mathbf{z}) = \exp\left(\mathbf{m}(\mathbf{x}_0)\cdot\mathbf{z}\right),$$

with $\mathbf{m}(\mathbf{x}_0) = (-x_{0,1}^2/2, \dots, -x_{0,n}^2/2)$. In this case, the expression (28) in Proposition 2.8 simplifies to:

$$\overline{m}(t) = \sum_{i=1}^{n}\left(\frac{x_{0,i}^2}{2}\left(\tanh^2(\mu_i t) - 1\right) - \frac{\mu_i}{2}\tanh(\mu_i t)\right). \tag{34}$$

Trajectory of adaptation in the presence of anisotropy: an illustrative example

We take $n = 3$ and $\mu_1 > \mu_2 > \mu_3$. The corresponding trajectory of $\overline{m}(t)$ is depicted in Figure 1. After a brief initial decay which was already observed in the isotropic case [32], $\overline{m}(t)$ rapidly increases and reaches a first plateau (of value close to $\overline{m}_{\infty,1} := -x_{0,2}^2/2 - x_{0,3}^2/2 - \mu_1/2$). Then, $\overline{m}(t)$ rapidly increases again to reach a second plateau (of value close to $\overline{m}_{\infty,2} := -x_{0,3}^2/2 - \mu_1/2 - \mu_2/2$). Finally, $\overline{m}(t)$ increases again and stabilises around $\overline{m}_\infty = -\mu_1/2 - \mu_2/2 - \mu_3/2$. Interestingly, although the ultimate value $\overline{m}_\infty$ does not depend on the initial phenotype, the intermediate plateaus depend on $\mathbf{x}_0$. Their values approximately correspond to the fitnesses associated with the successive projections of $\mathbf{x}_0$ on the hyperplanes $\{x_1 = 0\}$ and $\{x_1 = x_2 = 0\}$ minus the mutation load (we recall that the optimal phenotype was fixed at $\mathbf{x} = (0, \dots, 0)$, see Introduction).

More generally speaking, we prove in Section 4.5 that, for $n \geq 2$ and $\mu_1 > \dots > \mu_n$, the trajectory exhibits $(n-1)$ plateaus (before the final one of value $\overline{m}_\infty$), of respective values:

$$\overline{m} \approx \overline{m}_{\infty,k} := -\sum_{i=k+1}^{n}\frac{x_{0,i}^2}{2} - \sum_{i=1}^{k}\frac{\mu_i}{2}, \tag{35}$$

for $k = 1, \dots, (n-1)$. More precisely, we obtain the following proposition.

Proposition 2.13. *Let $\overline{m}(t)$ and $\overline{m}_{\infty,k}$ be defined by* (34) *and* (35)*, respectively. Given $\mathbf{x}_0 = (x_{0,1}, \dots, x_{0,n}) \in \mathbb{R}^n$, $T > 0$, $\varepsilon > 0$ and $\mu_1 > 0$, there exist $\mu_2, \dots, \mu_n > 0$ such that, for each $k \in [\![1, n-1]\!]$, the set:*

$$\left\{t \geq 0,\ |\overline{m}(t) - \overline{m}_{\infty,k}| \leq \varepsilon\right\},$$

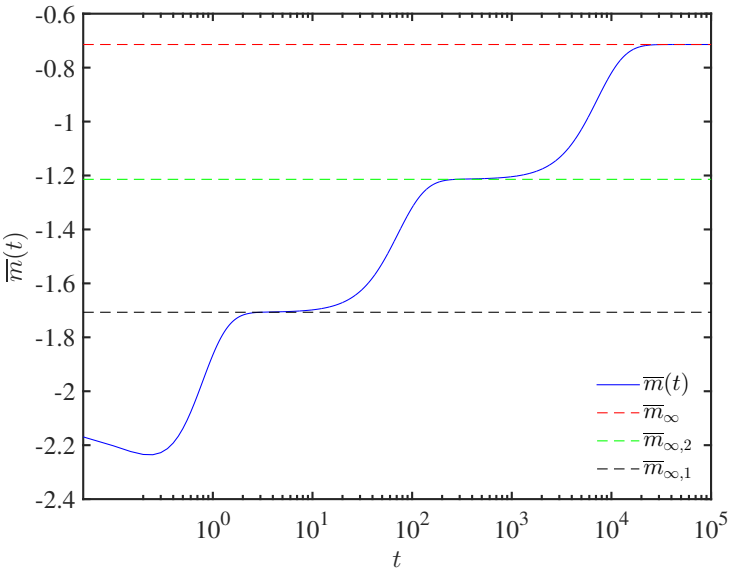


Figure 1 – **Trajectory of adaptation in the presence of anisotropy, with $n = 3$.** The function $\overline{m}(t)$ is given by formula (34), and the approximations of the values of the intermediate plateaus are $\overline{m}_{\infty,1} = -x_{0,2}^2/2 - x_{0,3}^2/2 - \mu_1/2$ and $\overline{m}_{\infty,2} = -x_{0,3}^2/2 - \mu_1/2 - \mu_2/2$. The mutational parameters are $\mu_1 = \sqrt{2}$, $\mu_2 = \sqrt{2} \cdot 10^{-2}$ and $\mu_3 = \sqrt{2} \cdot 10^{-4}$. The other parameter values are $\mathbf{x}_0 = (3/2, 1, 1)$.

contains an interval of length at least equal to T.

Note that the values $\overline{m}_{\infty,k}$ may not be ordered by increasing order, depending on the parameter values, possibly leading to nonmonotone trajectories of $\overline{m}(t)$. The plateaus are more visible when the μ_i's have different orders of magnitude. More precisely, we show in Section 4.5 that, given $T > 0$, $\overline{m}(t)$ remains around each plateau of value $\overline{m}_{\infty,k}$ at least during a time interval of duration T, for a good choice of the parameters μ_i.

Remark 2.14. *The proof of Proposition 2.13 does not lead to explicit expressions for the intervals at which the plateaus occur. In the particular case $n = 2$, we can obtain a more precise characterization of these intervals, see Section 4.5. In particular, taking $\mu_1 = 1$, $\mu_2 = 10^{-k}$ for some $k \geq 1$ and $x_{0,1} = x_{0,2} = 1$, we show that:*

(i) $\overline{m}'(t) \to -1/2 - 10^{-2k}/2$ as $t \to 0$;

(ii) at $t = t_0 = \ln(3)$, $\overline{m}'(t_0) > 27/250 - \mu_2/2 > 0.1 - 10^{-k}/2$;

(iii) for all t in the interval $I := \left(\ln\left(2 \cdot 10^k\right), 10^{k/2}\right)$, $\overline{m}'(t)$ is of order $10^{-3k/2}$;

(iv) at $t = t_1 = 10^k \ln[(\sqrt{2} + \sqrt{6})/2] > 10^{k/2}$, $\overline{m}'(t)$ is of order 10^{-k}.

At larger times, we already know that $\overline{m}'(t) \to 0$. Finally, this shows that $\overline{m}(t)$ remains stable within the interval I, corresponding to a part of the plateau, relatively to the times t_0 (before the interval I) and t_1 (after the interval I).

Long term evolution experiment with *Escherichia coli*

The long term evolution experiment (LTEE) has been carried by Lenski and his collaborators since 1988 [23]. Twelve populations of *E. coli* have been founded from a single common ancestor,

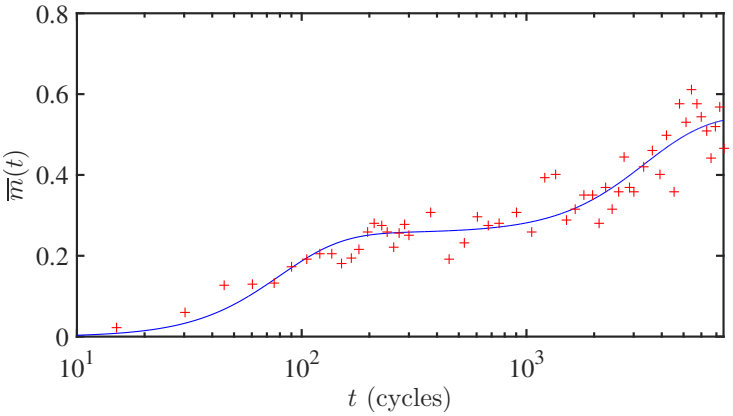
and are still evolving after more than 70,000 generations. The fitness evolved rapidly during the first 2,000 generations, and then remained nearly static between generations $5,000$ and $10,000$ [24], which would at least phenomenologically advocate for the existence of a phenotype optimum. However, more recent data (after generation 10,000) indicate that the mean fitness seems to increase without bounds [45]. Our goal here is not to propose a new explanation of the LTEE data, but simply to check whether the anisotropic model (4) leads to a better fit than an isotropic model.

The interpretation of the fitness data from the LTEE is quite subtle (see the comments in Richard Lenski's Experimental Evolution website http://myxo.css.msu.edu/ ecoli/srvsrf.html). For the sake of simplicity, as our objective is to check if the trajectories given by (34) can be qualitatively consistent with the fitness data from the LTEE, we make the following simplifying assumptions: (1) a time unit t in our model corresponds to 1 cycle of the experiment (1 day), corresponding to ≈ 6.64 generations [23]; (2) the link between $\overline{m}(t)$ and the mean fitness values given by the LTEE data, which are of the form $w(\text{gen})$ (gen is the generation number and w the Darwinian mean fitness, which is related to the Malthusian fitness through an exponential) is made by assuming that $\overline{m}(t+2) - \overline{m}(2) = \ln[w(6.64t)]$. Thus, we assume (arbitrarily) that 2 cycles were necessary to build the founding colony from the single ancestor (hence the term $t+2$). Additionally, in the data, the fitness is a relative fitness against the ancestor, which implies that $w(0) = 1$; this is why the quantity $\overline{m}(2)$ was subtracted to $\overline{m}(t+2)$. As mentioned above, the data are available for 12 populations. Here, we only use the data from one population (the population Ara-1, see [45]), for which measurements were obtained at 100-generation intervals during the first 2,000 generations, and then at 500-generation intervals.

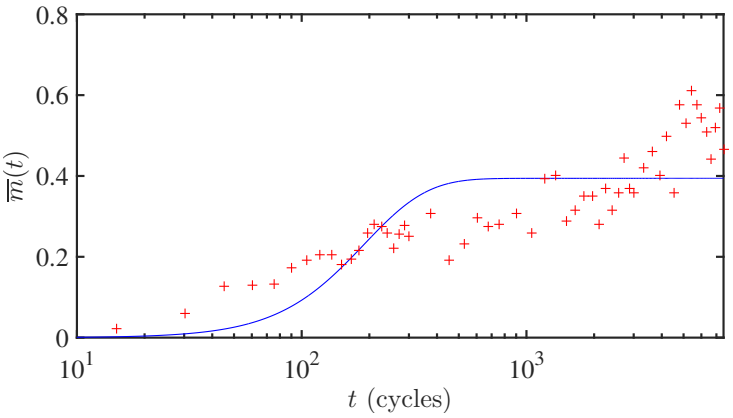
We carried out a fit (using Matlab® Curve Fitting Toolbox®, with a nonlinear least squares method) with the function $f(t) = \overline{m}(t+2) - \overline{m}(2)$. For the sake of simplicity, we assumed a two-dimensional phenotypic space ($n = 2$). The only parameters to estimate are μ_1, μ_2 and $\mathbf{x}_0$. We compared the result of this fit with a fit of the isotropic model ($\mu_1 = \mu_2$). The results are depicted in Figure 2. A graphical analysis shows that the anisotropic model gives a far better fit. This is confirmed by the adjusted R^2: 0.89 for the anisotropic model *versus* 0.57 for the isotropic model (R^2=1 indicates that the fitted model explains all of the variability in the data). In the anisotropic model, the fitted mutational parameters have two different orders of magnitude: $\mu_1 = 1.3 \cdot 10^{-2}$ and $\mu_2 = 3.0 \cdot 10^{-4}$. This leads to a first plateau until about 1,000 cycles ($6,640$ generations) followed by a second increase of the fitness. As expected, the isotropic model cannot explain this type of pattern.

3 Discussion

We considered a natural n-dimensional extension of the standard diffusive 'replicator-mutator' equation, to describe the dynamics of a phenotype distribution under anisotropic mutation and selection effects, and in the presence of a phenotype optimum. We proved that the corresponding Cauchy problem was well-posed (existence and uniqueness of the solution) and we proposed a new and general framework to the study of the quantitative behavior of the solution $q(t, \mathbf{x})$. This framework enabled us to derive a formula for the mean fitness in the population at all time



(a) Anisotropic model



(b) Isotropic model

Figure 2 – **Trajectory of adaptation, anisotropic and isotropic model *versus* LTEE data.** The function $f(t) = \overline{m}(t+2) - \overline{m}(2)$, with $\overline{m}(t)$ given by (34) was fitted to the LTEE data. The functions $f(t)$ with the parameter values corresponding to the best fit are depicted in blue. The red crosses correspond to the LTEE data (population Ara-1, $\ln[w(6.64t)]$, with w the original data). The values leading to the best fit are (a, anisotropic model) $\mathbf{x}_0 = (0.73, 0.76)$, $\mu_1 = 1.3 \cdot 10^{-2}$ and $\mu_2 = 3.0 \cdot 10^{-4}$; (b, isotropic model) $\mathbf{x}_0 = (0.30, 0.09)$ and $\mu_1 = \mu_2 = 5.3 \cdot 10^{-3}$.

(equation (28)), an important quantity to describe the dynamics of adaptation.

The case of an initially Gaussian distribution q_0 of the phenotypes is simpler, as the phenotype distribution $q(t, \mathbf{x})$ remains Gaussian (though anisotropic) at all time. However, when q_0 is not Gaussian, this result obviously breaks down. In this case, the method that we proposed consists in two steps: (1) to derive a degenerate nonlocal parabolic equation satisfied by the distribution $\mathbf{p}(t, \mathbf{m})$ of the fitness components, *i.e.*, the joint distribution of the $m_i = -x_i^2/2$. This equation is degenerate at 0, but has the advantage of being linear with respect to $\mathbf{m}$; (2) to derive a nonlocal transport equation satisfied by the cumulant generating function of $\mathbf{p}(t, \mathbf{m})$. This last equation can be solved analytically, and its solution leads to explicit formulae for all the moments of $\mathbf{p}$.

Conversely, the methods that are developed in this paper could be applied to solve more general degenerate parabolic PDEs. The idea would be firstly to transform the degenerate equation into a non-degenerate equation, through a change of function of the type (16) (rewriting $q^{\#}$ in terms of $\mathbf{p}$) to get an existence result, and secondly to consider PDEs satisfied by moment generating functions in order to obtain uniqueness and more quantitative properties of the solution of the degenerate PDE.

A natural idea to solve (4) could be to consider directly the cumulant generating function associated with q:

$$C_q(t, \mathbf{z}) = \ln\left(\int_{\mathbb{R}^n} e^{\mathbf{z}\cdot\mathbf{x}} q(t, \mathbf{x})\, d\mathbf{x}\right). \tag{36}$$

One would then have to see for which $\mathbf{z}$ this quantity makes sense, since now one integrates with respect to $\mathbf{x} \in \mathbb{R}^n$. Due to the nonlinear term $m(\mathbf{x})$ in (4), the equation satisfied by C_q would then be a nonlocal second-order viscous Hamilton-Jacobi type equation:

$$\partial_t C_q(t, \mathbf{z}) = -\frac{1}{2}\Delta C_q(t, \mathbf{z}) - \frac{1}{2}\|\nabla C_q(t, \mathbf{z})\|^2 + \sum_{i=1}^{n} \frac{\mu_i^2}{2} z_i^2 - \overline{m}(t),$$

whereas equation (23) for $C_{\mathbf{p}}$ was a nonlocal first-order transport equation.

From an applied perspective, the results of Section 2.3 illustrate the importance of taking anisotropy into account, as it can open up further explanations of experimental data. In sharp contrast with the known results based on isotropic models, our results show that the trajectory of adaptation may exhibit $(n-1)$ plateaus before it reaches the final one. In particular, using our analytic formulae for the dynamics of $\overline{m}(t)$, we obtained a very good fit of one of the most famous experimental dataset in experimental evolution, for which several evolutionary 'epochs' had already been observed [19]. This suggests that the FGM, in its anisotropic form, can generate a diversity of behaviours that may reconcile the various outcomes observed in long *versus* short term experimental evolution. However, whether this versatility may lead to overparameterization remains an open statistical question.

Model extensions

The simple formalism of model (4) can easily be extended to more general situations. A first generalization corresponds to the case where the environment changes over time, which would be described by a moving optimum changing smoothly with some external environmental factor

(*e.g.*, a drug dose, a temperature, *etc.*). Such a moving optimum could be taken into account by setting:

$$m(t,\mathbf{x}) = -\frac{\|\mathbf{x}-\mathbf{O}(t)\|^2}{2},$$

with $\mathbf{O}(t) \in \mathbb{R}^n$ the optimum at time t, instead of (5). In this case, we conjecture that Gaussian solutions of the form (12) still exist, suggesting that most our results could be extended to this situation.

Another generalization would be to explicitly consider birth-dependent mutation. Indeed, in many cases, mutations occur at birth. For example, in unicellulars undergoing a birth-death process, divisions without mutation occur at per capita rate $b\,(1-u)$ and divisions with mutation occur at rate $b\,u$ where u is a per division mutation probability. If the birth rate varies across genotypes or over time, then so does the mutation rate, and we can no longer consider a constant per capita mutation rate U. First, let us note that this complication can be ignored in several biologically relevant situations:

- In multicellulars, evolving over a discrete time demographic process (*e.g.* annual plants), the process of birth and mutations are decorrelated as what determines the mutation rate is the number of divisions in the germline, which does not a priori correlate with fitness (fecundity or survival). Hence the model with constant U can be applied here, our model being a continuous time approximation to the exact discrete time model.
- In many microbial evolution experiments (like the LTEE presented in Section 2.3), cells undergo rounds of a pure birth process separated by regular dilutions by a constant factor. In this situation, if cells grow to a constant density, then remain 'dormant' before dilution (so-called stationary phase in microbiology) then each cycle corresponds to a constant number of divisions. With a faster birth rate, these divisions simply occur earlier on during the cycle, but are not more numerous. This process becomes approximately equivalent to a discrete time life cycle at the scale of growth dilution cycles. The mutation rate per cycle is then also constant (independent of the birth rate). Note that this heuristics is only approximate, as it ignores variation in the birth rate between genotypes, relative to its mean, among co-segregating lineages, over one growth cycle. However, that this heuristics applies to the LTEE may explain why models that ignore the coupling of birth and mutation rate can provide a good fit to data from the LTEE. More detailed individual-based stochastic models for the LTEE can be found in *e.g.* [5, 7].
- Under 'viability selection', selection only acts to reduce the death rate d, while the birth rate b shows limited evolution: mutation rates are then also roughly constant over time ($U \approx \bar{b}\,u$, where $\bar{b}$ is the mean birth rate, stable over time). This happens if (i) the environmental challenge mainly affects the death rate and (ii) birth and death rates are uncorrelated so that evolution of d has little impact on b. In the context of phenotype-fitness landscape models, one situation where this applies is when fitness only depends on phenotype through the death rate, $m(\mathbf{x}) = b - d(\mathbf{x})$: the mutation rate is then $U = b\,u$, stable over time and across genotypes/phenotypes.

Overall there are several situations where one could ignore the connection between birth and mutation rates. Yet, in a more general context where the birth rate $b(\mathbf{x}) \geq 0$ does depend on phenotype $\mathbf{x}$, then the mutation rate *per capita* per unit time also depends on $\mathbf{x}$: $U(\mathbf{x}) = u\,b(\mathbf{x})$,

with u the probability of mutation per birth event. The model (4) can then be extended, to take into account this effect:

$$\partial_t q(t,\mathbf{x}) = \sum_{i=1}^{n} \frac{1}{2}\,\partial_{ii}[\mu_i^2(\mathbf{x})q(t,\mathbf{x})] + (m(\mathbf{x}) - \overline{m}(t))\,q(t,\mathbf{x}),\ t>0,\ \mathbf{x}\in\mathbb{R}^n,$$

with $\mu_i^2(\mathbf{x}) = \lambda_i\,U(\mathbf{x}) = \lambda_i\,u\,b(\mathbf{x})$, with λ_i the mutational variance at each trait, see Appendix A. However, the mathematical results of our paper cannot be straightforwardly extended to this case, even in the isotropic case. In particular, as soon as $b(\mathbf{x})$ is not constant, this equation has no Gaussian solution. Still, characterization of the steady states should be possible as a principal eigenvalue problem.

4 Proofs

This section is devoted to the proofs of the results announced in Section 2. Section 4.1 is concerned with problem (10) satisfied by the distribution of phenotype frequencies $q(t,\mathbf{x})$, while Section 4.2 deals with the fitness frequencies $\mathbf{p}(t,\mathbf{m})$ and the proofs of Proposition 2.3 and Theorem 2.4. In Section 4.3, we carry out the proofs of Theorem 2.6 and Proposition 2.8 on the cumulant generating functions, and their corollaries (Theorem 2.5 and Corollaries 2.7 and 2.9) in the isotropic case. Lastly, Sections 4.4 and 4.5 are concerned with the stationary states and the existence of plateaus for the mean fitness.

4.1 Proofs of Theorem 2.1 and Corollary 2.2 on the Cauchy problem (10)

Before considering the nonlocal problem (10), we begin with the local problem:

$$\left\{\begin{array}{rcl} \partial_t v(t,\mathbf{x}) &=& \displaystyle\sum_{i=1}^{n}\frac{\mu_i^2}{2}\partial_{ii}v(t,\mathbf{x}) + m(\mathbf{x})\,v(t,\mathbf{x}),\ t\geq 0,\ \mathbf{x}\in\mathbb{R}^n,\\ v(0,\mathbf{x}) &=& q_0(\mathbf{x}),\ \mathbf{x}\in\mathbb{R}^n, \end{array}\right. \tag{37}$$

where q_0 satisfies (7)-(9). As the fitness function:

$$m(\mathbf{x}) = -\frac{\|\mathbf{x}\|^2}{2},$$

is unbounded, standard parabolic theory with bounded coefficients does not apply. However, some properties of (37) can be for instance be found in [4, 26, 36], which in particular lead to the following result.

Theorem 4.1. *The problem* (37) *admits a unique bounded solution* $v\in C^{1,2}(\mathbb{R}_+\times\mathbb{R}^n)$. *Additionally, we have:*

$$\forall T>0,\ \exists\, S>0,\ \forall\, t\in[0,T],\quad \|v(t,\cdot)\|_{C^{2+\alpha}(\mathbb{R}^n)} \leq S\,\|q_0\|_{C^{2+\alpha}(\mathbb{R}^n)}, \tag{38}$$

and:

$$0 < v(t,\mathbf{x}) < (K_t * q_0)(\mathbf{x}),\ \textit{for all}\ (t,\mathbf{x})\in\mathbb{R}_+^*\times\mathbb{R}^n, \tag{39}$$

with:

$$K_t(\mathbf{x}) = \frac{1}{(2\pi t)^{n/2}\,\mu_1\cdots\mu_n}\,\exp\left[-\sum_{i=1}^{n}\frac{x_i^2}{2\mu_i^2 t}\right]. \tag{40}$$

Proof. Let us fix a time $T > 0$. Theorem 2 in [26] implies that (37) admits a unique bounded solution $v \in C^{1,2}([0,T] \times \mathbb{R}^n)$ and this solution satisfies (38). Theorem III in [4] further implies that this solution is nonnegative. As T was chosen arbitrarily, these existence, uniqueness and nonnegativity results extend to $t \in (0,+\infty)$, with local boundedness in t.

Let us set $h(t,\mathbf{x}) := (K_t * q_0)(\mathbf{x})$ for $t > 0$, and $h(0,\mathbf{x}) = q_0(\mathbf{x})$. The function h satisfies:

$$\partial_t h(t,\mathbf{x}) = \sum_{i=1}^n \frac{\mu_i^2}{2} \partial_{ii} h(t,\mathbf{x}),$$

for all $t > 0$ and $\mathbf{x} \in \mathbb{R}^n$. Let $\psi(t,\mathbf{x}) := v(t,\mathbf{x}) - h(t,\mathbf{x})$. We see that, for all $t > 0$ and $\mathbf{x} \in \mathbb{R}^n$,

$$\partial_t \psi(t,\mathbf{x}) - \sum_{i=1}^n \frac{\mu_i^2}{2} \partial_{ii} \psi(t,\mathbf{x}) = m(\mathbf{x}) v(t,\mathbf{x}) \leq 0, \tag{41}$$

and $\psi(0,\mathbf{x}) = 0$. By the Phragmèn-Linderlöf principle [36, Theorem 10, Chapter 3], we get that $\psi \leq 0$ in $\mathbb{R}_+ \times \mathbb{R}^n$, *i.e.*, $v(t,\mathbf{x}) \leq h(t,\mathbf{x}) = (K_t * q_0)(\mathbf{x})$. We therefore infer that $0 \leq v(t,\mathbf{x}) \leq (K_t * q_0)(\mathbf{x})$ in $(0,+\infty) \times \mathbb{R}^n$. Since q_0 is bounded, this also implies that v is bounded in $\mathbb{R}_+ \times \mathbb{R}^n$.

By the standard strong parabolic maximum principle, we conclude that the first inequality is strict, *i.e.*, $0 < v(t,\mathbf{x})$, for all $(t,\mathbf{x}) \in (0,+\infty) \times \mathbb{R}^n$, since $v(0,\cdot) = q_0$ is continuous, nonnegative and not identically equal to 0. Furthermore, since the inequality in (41) is then strict for all $(t,\mathbf{x}) \in (0,+\infty) \times (\mathbb{R}^n \setminus \{\mathbf{O}\})$, we get that $\psi(t,\mathbf{x}) < 0$, *i.e.*, $v(t,\mathbf{x}) < (K_t * q_0)(\mathbf{x})$, for all $(t,\mathbf{x}) \in (0,+\infty) \times \mathbb{R}^n$. □

In order to connect (37) and (10), we need the following lemma.

Lemma 4.2. The function:

$$t \mapsto \overline{m}_v(t) := \int_{\mathbb{R}^n} m(\mathbf{x})\, v(t,\mathbf{x})\, d\mathbf{x}, \tag{42}$$

is real-valued and continuous in $\mathbb{R}_+$ and, for every $t \geq 0$, there holds:

$$1 + \int_0^t \overline{m}_v(s)\, ds = 1 + \int_0^t \int_{\mathbb{R}^n} m(\mathbf{x}) v(s,\mathbf{x})\, d\mathbf{x}\, ds = \int_{\mathbb{R}^n} v(t,\mathbf{x})\, d\mathbf{x} > 0. \tag{43}$$

Proof. First of all, denote:

$$\underline{\mu} := \min(\mu_1, \ldots, \mu_n) > 0 \quad \text{and} \quad \overline{\mu} := \max(\mu_1, \ldots, \mu_n) > 0. \tag{44}$$

It follows from the assumptions on q_0 that $\overline{m}_v(0)$ is a nonpositive real number. Consider now any $t > 0$ and let us check that $\overline{m}_v(t)$ defined in (42) is a nonpositive real number. The function $\mathbf{x} \mapsto m(\mathbf{x})\, v(t,\mathbf{x})$ is nonpositive and continuous in $\mathbb{R}^n$. Furthermore, it follows from Theorem 4.1 and the assumptions on q_0 that:

$$\begin{aligned}
\int_{\mathbb{R}^n} |m(\mathbf{x})|\, v(t,\mathbf{x})\, d\mathbf{x} &\leq \frac{1}{2(2\pi t)^{n/2} \underline{\mu}^n} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \exp\left(-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2\overline{\mu}^2 t}\right) g(\|\mathbf{y}\|)\, \|\mathbf{x}\|^2\, d\mathbf{y}\, d\mathbf{x} \\
&= \frac{\overline{\mu}^n}{2\pi^{n/2} \underline{\mu}^n} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \exp(-\|\mathbf{z}\|^2)\, g(\|\mathbf{x} - \overline{\mu}\sqrt{2t}\mathbf{z}\|)\, \|\mathbf{x}\|^2\, d\mathbf{z}\, d\mathbf{x} \\
&\leq \frac{\overline{\mu}^n}{2\pi^{n/2} \underline{\mu}^n} \int_{\mathbb{R}^n} \left(\int_{\|\mathbf{z}\| \leq \|\mathbf{x}\|/(2\overline{\mu}\sqrt{2t})} \exp(-\|\mathbf{z}\|^2)\, g(\|\mathbf{x}\|/2)\, d\mathbf{z} \right. \\
&\qquad \left. + \int_{\|\mathbf{z}\| > \|\mathbf{x}\|/(2\overline{\mu}\sqrt{2t})} \exp(-\|\mathbf{z}\|^2)\, g(0)\, d\mathbf{z} \right) \|\mathbf{x}\|^2\, d\mathbf{x} \\
&\leq \frac{\overline{\mu}^n}{2\pi^{n/2} \underline{\mu}^n} \left(\pi^{n/2} \int_{\mathbb{R}^n} g(\|\mathbf{x}\|/2)\, \|\mathbf{x}\|^2\, d\mathbf{x} + g(0) \int_{\mathbb{R}^n} h_t(\|\mathbf{x}\|)\, \|\mathbf{x}\|^2\, d\mathbf{x} \right),
\end{aligned}$$

where:

$$h_t(r) := \int_{\|\mathbf{z}\| \geq r/(2\overline{\mu}\sqrt{2t})} \exp(-\|\mathbf{z}\|^2)\, d\mathbf{z} = O(\exp(-r)), \quad \text{as } r \to +\infty.$$

Since the integral $\int_{\mathbb{R}^n} \|\mathbf{x}\|^2 g(\|\mathbf{x}\|)\, d\mathbf{x}$ converges by assumption, one concludes that the integral $\int_{\mathbb{R}^n} |m(\mathbf{x})|\, v(t,\mathbf{x})\, d\mathbf{x}$ converges as well, hence $\overline{m}_v(t)$ is a nonpositive real number. Furthermore, since the quantities $h_t(r)$ are non-decreasing with respect to $t > 0$, the same arguments together with Lebesgue's dominated convergence theorem imply straightforwardly that the function $\overline{m}_v$ is continuous in $\mathbb{R}_+$.

The convergence of the integral defining $\overline{m}_v(t)$ for every $t \geq 0$, together with the nonnegativity and continuity of v and $\mathbf{x} \mapsto -m(\mathbf{x}) = \|\mathbf{x}\|^2/2$, immediately implies that the integral $\int_{\mathbb{R}^n} v(t,\mathbf{x})\, d\mathbf{x}$ is a nonnegative real number for each $t \geq 0$. Moreover, since $v(t,\mathbf{x}) > 0$ for all $(t,\mathbf{x}) \in \mathbb{R}_+^* \times \mathbb{R}^n$, we infer that:

$$\int_{\mathbb{R}^n} v(t,\mathbf{x})\, d\mathbf{x} > 0 \quad \text{for all } t > 0.$$

As in the previous paragraph, the function $t \mapsto \int_{\mathbb{R}^n} v(t,\mathbf{x})\, d\mathbf{x}$ is also continuous in $\mathbb{R}_+$.

Fix now an arbitrary $t > 0$. For $R > 0$, denote $B_R = \{\mathbf{x} \in \mathbb{R}^n,\ \|\mathbf{x}\| < R\}$, ν the outward unit normal on ∂B_R and $d\sigma(\mathbf{x})$ the surface measure on ∂B_R. For every $\varepsilon \in (0,t)$ and every $R > 0$, the integration of (37) over $[\varepsilon, t] \times B_R$ yields:

$$\int_{B_R} v(t,\mathbf{x})\, d\mathbf{x} - \int_{B_R} v(\varepsilon,\mathbf{x})\, d\mathbf{x} = \frac{1}{2} \int_\varepsilon^t \int_{\partial B_R} \sum_{i=1}^n \mu_i^2\, \nu_i\, \partial_i v(s,\mathbf{x})\, d\sigma(\mathbf{x})\, ds + \int_\varepsilon^t \int_{B_R} m(\mathbf{x})\, v(s,\mathbf{x})\, d\mathbf{x}\, ds.$$

Since the first term of the right-hand side converges to 0 as $R \to +\infty$ from standard parabolic estimates (see [14]), one gets that:

$$\int_{\mathbb{R}^n} v(t,\mathbf{x})\, d\mathbf{x} - \int_{\mathbb{R}^n} v(\varepsilon,\mathbf{x})\, d\mathbf{x} = \int_\varepsilon^t \overline{m}_v(s)\, ds,$$

by passing to the limit $R \to +\infty$. The limit $\varepsilon \to 0^+$ then yields:

$$\int_{\mathbb{R}^n} v(t,\mathbf{x})\, d\mathbf{x} - 1 = \int_{\mathbb{R}^n} v(t,\mathbf{x})\, dx - \int_{\mathbb{R}^n} q_0(\mathbf{x})\, d\mathbf{x} = \int_0^t \overline{m}_v(s)\, ds, \tag{45}$$

which gives the desired result (43) for $t > 0$. Formula (43) for $t = 0$ is trivial since $v(0,\cdot) = q_0$ has unit mass. The proof of Lemma 4.2 is thereby complete. □

Proof of Theorem 2.1. Let $v \geq 0$ be the unique classical solution of (37) given in Theorem 4.1. From Lemma 4.2, the function $\overline{m}_v$ defined by (42) is continuous in $\mathbb{R}_+$. Let us then set:

$$q(t,\mathbf{x}) = \frac{v(t,\mathbf{x})}{1 + \int_0^t \overline{m}_v(s)\, ds}, \tag{46}$$

for every $(t,\mathbf{x}) \in \mathbb{R}_+ \times \mathbb{R}^n$. We recall that the denominator in the right-hand side of (46) is positive from Lemma 4.2. From Theorem 4.1 and Lemma 4.2, the function q is nonnegative and of class $C^{1,2}(\mathbb{R}_+ \times \mathbb{R}^n)$. Furthermore, $q(0,\cdot) = v(0,\cdot) = q_0$ in $\mathbb{R}^n$ and it follows from Lemma 4.2 that:

$$\forall\, t \geq 0, \quad \int_{\mathbb{R}^n} q(t,\mathbf{x})\, d\mathbf{x} = 1, \qquad \text{and} \qquad \overline{m}(t) := \int_{\mathbb{R}^n} m(\mathbf{x})\, q(t,\mathbf{x})\, d\mathbf{x} = \frac{\overline{m}_v(t)}{1 + \int_0^t \overline{m}_v(s)\, ds},$$

hence the function $\overline{m}$ is real-valued, nonpositive and continuous in $\mathbb{R}_+$. Lastly, it is immediate to see that q obeys (10). Furthermore, since v is bounded in $(0,T)\times\mathbb{R}^n$, the function q is bounded in $(0,T)\times\mathbb{R}^n$ for every $T>0$.

To show the uniqueness, assume now that we have two nonnegative classical solutions q_1 and q_2 of (10) in $C^{1,2}(\mathbb{R}_+\times\mathbb{R}^n)\cap L^\infty((0,T)\times\mathbb{R}^n)$ (for every $T>0$), with the same initial condition q_0 satisfying (7)-(9), and such that the functions:

$$\overline{m}_1(t) := \int_{\mathbb{R}^n} m(\mathbf{x})\, q_1(t,\mathbf{x})\, d\mathbf{x}, \quad \text{and} \quad \overline{m}_2(t) := \int_{\mathbb{R}^n} m(\mathbf{x})\, q_1(t,\mathbf{x})\, d\mathbf{x},$$

are real-valued and continuous in $\mathbb{R}_+$. Define:

$$v_i(t,\mathbf{x}) = q_i(t,\mathbf{x})\, \exp\Big(\int_0^t \overline{m}_i(s)ds\Big),$$

for $i=1,2$, and $(t,\mathbf{x})\in\mathbb{R}_+\times\mathbb{R}^n$. The two functions v_1 and v_2 satisfy (37) and are bounded in $(0,T)\times\mathbb{R}^n$ for every $T>0$. From Theorem 4.1, we get $v_1\equiv v_2$ in $\mathbb{R}_+\times\mathbb{R}^n$. Furthermore, for all $i=1,2$ and $t\in\mathbb{R}_+$, there holds:

$$\overline{m}_{v_i}(t) = \int_{\mathbb{R}^n} m(\mathbf{x})\, v_i(t,\mathbf{x})\, d\mathbf{x} = \overline{m}_i(t)\, \exp\left(\int_0^t \overline{m}_i(s)ds\right) = \frac{d}{dt}\left[\exp\left(\int_0^t \overline{m}_i(s)ds\right)\right].$$

Hence, for all $(t,\mathbf{x})\in\mathbb{R}_+\times\mathbb{R}^n$, we get:

$$q_1(t,\mathbf{x}) = \frac{v_1(t,\mathbf{x})}{1+\int_0^t \overline{m}_{v_1}(s)\, ds} = \frac{v_2(t,\mathbf{x})}{1+\int_0^t \overline{m}_{v_2}(s)\, ds} = q_2(t,\mathbf{x}).$$

The proof of Theorem 2.1 is thereby complete. □

Proof of Corollary 2.2. It is a straightforward calculation to check that the function q defined by (12)-(13) is a classical bounded solution of (10) with initial condition given by (11). The conclusion then follows from the uniqueness part of Theorem 2.1. □

4.2 A degenerate parabolic PDE satisfied by $\mathbf{p}(t,\mathbf{m})$

Proof of Proposition 2.3. We recall that the phenotypes are represented by n traits, and so by a vector in $\mathbb{R}^n$ and that we have a constant optimum $\mathbf{O}$, which is fixed to $(0,\dots,0)$ up to translation. We define, for each $\varepsilon=(\varepsilon_1,\dots,\varepsilon_n)\in\{\pm1\}^n$, the subset:

$$Q_\varepsilon = \{\mathbf{x}\in\varepsilon_1\mathbb{R}_+^*\times\cdots\times\varepsilon_n\mathbb{R}_+^*\}\subset\mathbb{R}^n.$$

For any time $t\geq 0$, we get from the law of total probability that:

$$\int_{\mathbb{R}^n_-} \mathbf{p}(t,\mathbf{m})\, d\mathbf{m} = \int_{\mathbb{R}^n} q(t,\mathbf{x})\, d\mathbf{x} = 1,$$

and that, for any $m\in\mathbb{R}^n_-$,

$$\mathbf{p}(t,\mathbf{m}) = \sum_{\varepsilon\in\{\pm1\}^n} \mathbf{p}(t,\mathbf{m}\mid \mathbf{x}(\mathbf{m})\in Q_\varepsilon)\ \int_{Q_\varepsilon} q(t,\mathbf{y})d\mathbf{y}, \tag{47}$$

with $\mathbf{p}(t,\mathbf{m}\,|\,\mathbf{x}(\mathbf{m})\in Q_\varepsilon)$ the conditional density of the fitness vector $\mathbf{m}$, given that the associated phenotype $\mathbf{x}(\mathbf{m})$ is in Q_ε. In the above formula (47), we also use the fact that $\int_H q(t,\mathbf{x})\,d\mathbf{x}=0$ with:

$$H=\bigcup_{1\le i\le n}\{\mathbf{x}\in\mathbb{R}^n,\ x_i=0\},$$

since $q(t,\cdot)$ is continuous in $\mathbb{R}^n$.

As the fitness function $\mathbf{x}\in Q_\varepsilon\mapsto\mathbf{m}(\mathbf{x})=(m_1(\mathbf{x}),\dots,m_n(\mathbf{x}))$ with $m_i(\mathbf{x})=-x_i^2/2$ is one-to-one from Q_ε to $(\mathbb{R}_-^*)^n$, with inverse $\mathbf{m}\mapsto\mathbf{x}^\varepsilon(\mathbf{m})=(\varepsilon_1\sqrt{-2m_1},\dots,\varepsilon_n\sqrt{-2m_n})$, we infer that, for every $t\ge 0$ and $\mathbf{m}\in(\mathbb{R}_-^*)^n$,

$$\mathbf{p}(t,\mathbf{m}\mid\mathbf{x}(\mathbf{m})\in Q_\varepsilon)\ \int_{Q_\varepsilon}q(t,\mathbf{y})d\mathbf{y}=\frac{q(t,\mathbf{x}^\varepsilon(\mathbf{m}))}{|\det J_\varepsilon|}=\frac{2^{-n/2}}{\sqrt{|m_1\cdots m_n|}}\,q(t,\mathbf{x}^\varepsilon(\mathbf{m})),$$

with $J_\varepsilon=\mathrm{diag}(-\varepsilon_1\sqrt{-2m_1},\dots,-\varepsilon_n\sqrt{-2m_n})$. Finally, we get:

$$\mathbf{p}(t,\mathbf{m})=\frac{2^{-n/2}}{\sqrt{|m_1\cdots m_n|}}\sum_{\varepsilon\in\{\pm1\}^n}q(t,\mathbf{x}^\varepsilon(\mathbf{m})),$$

and:

$$\begin{aligned}\overline{m}(t)&=\sum_{i=1}^n\int_{\mathbb{R}^n}-\frac{x_i^2}{2}q(t,\mathbf{x})d\mathbf{x}=\sum_{i=1}^n\int_{\mathbb{R}^n\setminus H}-\frac{x_i^2}{2}q(t,\mathbf{x})d\mathbf{x},\\&=\sum_{i=1}^n\sum_{\varepsilon\in\{\pm1\}^n}\int_{\{\varepsilon_1x_1>0,\dots,\varepsilon_nx_n>0\}}-\frac{x_i^2}{2}\,q(t,\mathbf{x})d\mathbf{x},\\&=\sum_{i=1}^n\sum_{\varepsilon\in\{\pm1\}^n}\int_{(\mathbb{R}_-^*)^n}m_iq(t,\mathbf{x}^\varepsilon(\mathbf{m}))\frac{2^{-n/2}}{\sqrt{|m_1\cdots m_n|}}d\mathbf{m}=\sum_{i=1}^n\int_{(\mathbb{R}_-^*)^n}m_i\mathbf{p}(t,\mathbf{m})d\mathbf{m}.\end{aligned}$$

Notice that all integrals in the last sum converge since all integrands are nonpositive and the sum of these integrals is a real number. Observe lastly that $\int_{(\mathbb{R}_-^*)^n}m_i\mathbf{p}(t,\mathbf{m})d\mathbf{m}=\int_{\mathbb{R}_-^n}m_i\mathbf{p}(t,\mathbf{m})d\mathbf{m}$, for every $1\le i\le n$, since $\mathbf{p}(t,\cdot)$ is an $L^1(\mathbb{R}_-^n)$ function. The proof of Proposition 2.3 is thereby complete. □

Proof of Theorem 2.4. Formula (16) implies that the function $\mathbf{p}$ is of class $C^{1,2}(\mathbb{R}_+\times(\mathbb{R}_-^*)^n)$ with initial condition $\mathbf{p}_0$ given by (18). Furthermore, it is straightforward to check that, for all $t\ge 0$ and $\mathbf{m}\in(\mathbb{R}_-^*)^n$:

$$\begin{aligned}2^{-n/2}\ \partial_i\mathbf{p}(t,\mathbf{m})&=-\ \frac{q^{\#}(t,\mathbf{x}^{\mathbb{1}}(\mathbf{m}))}{2m_i\sqrt{|m_1\cdots m_n|}}-\ \frac{1}{\sqrt{2|m_i|}}\ \frac{\partial_iq^{\#}(t,\mathbf{x}^{\mathbb{1}}(\mathbf{m}))}{\sqrt{|m_1\cdots m_n|}},\\2^{-n/2}\ \partial_{ii}\mathbf{p}(t,\mathbf{m})&=\frac{3}{4}\ \frac{q^{\#}(t,\mathbf{x}^{\mathbb{1}}(\mathbf{m}))}{m_i^2\sqrt{|m_1\cdots m_n|}}+\frac{3}{2}\ \frac{\partial_iq^{\#}(t,\mathbf{x}^{\mathbb{1}}(\mathbf{m}))}{m_i\sqrt{2|m_i|}\sqrt{|m_1\cdots m_n|}}-\frac{1}{2m_i}\ \frac{\partial_{ii}q^{\#}(t,\mathbf{x}^{\mathbb{1}}(\mathbf{m}))}{\sqrt{|m_1\cdots m_n|}},\end{aligned}$$

with $\mathbf{x}^{\mathbf{1}}(\mathbf{m}) = (\sqrt{-2m_1}, \dots, \sqrt{-2m_n}) \in (\mathbb{R}_+^*)^n$. Hence, we have:

$$\begin{aligned}
&\sum_{i=1}^n \mu_i^2 \Big(m_i \partial_{ii} \mathbf{p}(t,\mathbf{m}) + \frac{3}{2} \partial_i \mathbf{p}(t,\mathbf{m}) \Big) \\
&\quad = - \sum_{i=1}^n \frac{2^{n/2}}{\sqrt{|m_1 \cdots m_n|}} \, \frac{\mu_i^2}{2} \partial_{ii} q^{\#}(t, \mathbf{x}^{\mathbf{1}}(\mathbf{m})), \\
&\quad = -2^{n/2} \, \frac{\partial_t q^{\#}(t, \mathbf{x}^{\mathbf{1}}(\mathbf{m})) - (m(\mathbf{x}^{\mathbf{1}}(\mathbf{m})) - \overline{m}(t)) \; q^{\#}(t, \mathbf{x}^{\mathbf{1}}(\mathbf{m}))}{\sqrt{|m_1 \cdots m_n|}}, \\
&\quad = -\partial_t \mathbf{p}(t,\mathbf{m}) + \Big(\sum_{i=1}^n m_i - \overline{m}(t) \Big) \; \mathbf{p}(t,\mathbf{m}).
\end{aligned}$$

Theorem 2.4 is thereby proven. □

Proof of Theorem 2.5. Since by definition $p(t,m)\,dm$ is the pushforward measure, at each time $t \geq 0$, of the measure $q(t,\mathbf{x})\,d\mathbf{x}$ by the map $\mathbf{x} \mapsto -\|\mathbf{x}\|^2/2$, it follows from the layer-cake formula that:

$$p(t,m) = (2|m|)^{n/2-1} \int_{\mathbb{S}^{n-1}} q\big(t, \sqrt{2|m|}\,\sigma\big)\, d\sigma = (2|m|)^{n/2-1}\, Q\big(t, \sqrt{2|m|}\big),$$

for all $(t,m) \in \mathbb{R}_+ \times \mathbb{R}_+^*$, where $\mathbb{S}^{n-1}$ denotes the unit Euclidean sphere of $\mathbb{R}^n$ and:

$$Q(t,r) = \int_{\mathbb{S}^{n-1}} q(t, r\,\sigma)\, d\sigma,$$

for $(t,r) \in \mathbb{R}_+ \times \mathbb{R}_+$. Since q is of class $C^{1,2}(\mathbb{R}_+ \times \mathbb{R}^n)$, it is easy to see that the function:

$$(t,\mathbf{x}) \mapsto \widetilde{q}(t,\mathbf{x}) := \int_{\mathbb{S}^{n-1}} q(t, \|x\|\,\sigma)\, d\sigma,$$

is of class $C^{1,2}(\mathbb{R}_+ \times \mathbb{R}^n)$ too, hence the function Q is of class $C^{1,2}(\mathbb{R}_+ \times \mathbb{R}_+)$ and p is of class $C^{1,2}(\mathbb{R}_+ \times \mathbb{R}_-^*)$. Furthermore, $\widetilde{q}(0,\cdot)$ is of class $C^{2+\alpha}(\mathbb{R}^n)$ since $q(0,\cdot) = q_0$ is of class $C^{2+\alpha}(\mathbb{R}^n)$.

Since q solves (10), which is invariant by rotation in the present isotropic case ($\mu_i = \mu$ for every $1 \leq i \leq n$) and since $\overline{m}(t) = \int_{\mathbb{R}^n} m(\mathbf{x})\, q(t,\mathbf{x})\, d\mathbf{x} = \int_{\mathbb{R}^n} m(\mathbf{x})\, \widetilde{q}(t,\mathbf{x})\, d\mathbf{x}$, for every $t \geq 0$, it follows that $\widetilde{q}$ solves (10) as well, with initial condition $\widetilde{q}(0,\cdot)$. As a consequence, Q satisfies:

$$\partial_t Q(t,r) = \frac{\mu^2}{2} \Big(\partial_{rr} Q(t,r) + \frac{n-1}{r} \partial_r Q(t,r) \Big) + \Big(-\frac{r^2}{2} - \overline{m}(t) \Big) Q(t,r),$$

for all $(t,r) \in \mathbb{R}_+ \times \mathbb{R}_+^*$. But since:

$$Q(t,r) = r^{2-n}\, p\Big(t, -\frac{r^2}{2}\Big),$$

for all $(t,r) \in \mathbb{R}_+ \times \mathbb{R}_+^*$, it is then straightforward to check that p solves (19) in $\mathbb{R}_+ \times \mathbb{R}_-^*$. Lastly, the formula (20) is an immediate consequence of the definitions of p and $\mathbf{p}$, and the proof of Theorem 2.5 is thereby complete. □

4.3 Generating functions

Proof of Theorem 2.6. Given q_0 satisfying (7)-(9), we call v the unique bounded nonnegative $C^{1,2}(\mathbb{R}_+ \times \mathbb{R}^n)$ solution of (37) defined in Theorem 4.1 with initial condition $q_0^{\#}$. Notice that the function $q_0^{\#}$ satisfies the same assumptions (7)-(9) as q_0. By uniqueness and symmetry

of (37) with respect to the change of variable x_i into $-x_i$, for any $1 \le i \le n$, it follows that $v(t,\mathbf{x}) = v^{\#}(t,\mathbf{x})$ and, as in (46),

$$q^{\#}(t,\mathbf{x}) = \frac{v^{\#}(t,\mathbf{x})}{1+\int_0^t \overline{m}_v(s)\,ds} = \frac{v(t,\mathbf{x})}{1+\int_0^t \overline{m}_v(s)\,ds}, \tag{48}$$

for all $(t,\mathbf{x}) \in \mathbb{R}_+ \times \mathbb{R}_+^n$.

As already noticed in Section 2.1, from (15) and the nonnegativity of $\mathbf{p}$, the functions $(t,\mathbf{z}) \mapsto M_{\mathbf{p}}(t,\mathbf{z})$ and $(t,\mathbf{z}) \mapsto C_{\mathbf{p}}(t,\mathbf{z})$ given in (21)-(22) are well defined in $\mathbb{R}_+ \times \mathbb{R}_+^n$. Let us start by proving the continuity of these functions $M_{\mathbf{p}}$ and $C_{\mathbf{p}}$ in $\mathbb{R}_+ \times \mathbb{R}_+^n$. Owing to the relations (16) and (48), we see that:

$$M_{\mathbf{p}}(t,\mathbf{z}) = \frac{2^{n/2}}{1+\int_0^t \overline{m}_v(s)ds} \int_{\mathbb{R}_-^n} \frac{e^{\mathbf{z}\cdot\mathbf{m}}}{\sqrt{|m_1\cdots m_n|}}\, v(t,\mathbf{x}^{\mathbf{1}}(\mathbf{m}))\,d\mathbf{m}, \tag{49}$$

for all $(t,\mathbf{z}) \in \mathbb{R}_+ \times \mathbb{R}_+^n$, with $\mathbf{x}^{\mathbf{1}}(\mathbf{m}) = (\sqrt{-2m_1},\ldots,\sqrt{-2m_n})$. Hence, we have:

$$\begin{aligned} M_{\mathbf{p}}(t,\mathbf{z}) &= \frac{2^n}{1+\int_0^t \overline{m}_v(s)ds} \int_{\mathbb{R}_+^n} \exp\Big(-\sum_{i=1}^n \frac{z_i x_i^2}{2}\Big)\, v(t,\mathbf{x})\,d\mathbf{x}, \\ &= \frac{1}{1+\int_0^t \overline{m}_v(s)ds} \int_{\mathbb{R}^n} \exp\Big(-\sum_{i=1}^n \frac{z_i x_i^2}{2}\Big)\, v(t,\mathbf{x})\,d\mathbf{x}. \end{aligned} \tag{50}$$

Notice that the function $t \mapsto \int_0^t \overline{m}_v(s)ds$ is continuous in $\mathbb{R}_+$. Furthermore, the function $(t,\mathbf{z}) \mapsto \exp(-\sum_{i=1}^n z_i x_i^2/2)\, v(t,\mathbf{x})$ is also continuous in $\mathbb{R}_+ \times \mathbb{R}_+^n$, for every $\mathbf{x} \in \mathbb{R}^n$. Lastly, as in the proof of Lemma 4.2, it follows from (9), (39)-(40) and (44) that, for any $\mathbf{z} \in \mathbb{R}_+^n$, $T > 0$, $t \in (0,T]$ and $\mathbf{x} \in \mathbb{R}^n$, there holds:

$$\begin{aligned} 0 &\le \exp\Big(-\sum_{i=1}^n \frac{z_i x_i^2}{2}\Big)\, v(t,\mathbf{x}) \le v(t,\mathbf{x}), \\ &\le \frac{1}{(2\pi t)^{n/2}\underline{\mu}^n} \int_{\mathbb{R}^n} \exp\Big(-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2\overline{\mu}^2 t}\Big)\, q_0^{\#}(\mathbf{y})\,d\mathbf{y}, \\ &\le \frac{\overline{\mu}^n}{\pi^{n/2}\underline{\mu}^n} \int_{\mathbb{R}^n} e^{-\|\mathbf{y}'\|^2}\, g(\|\mathbf{x}-\overline{\mu}\sqrt{2t}\mathbf{y}'\|)\,d\mathbf{y}', \\ &\le \frac{\overline{\mu}^n}{\pi^{n/2}\underline{\mu}^n}\Big[\int_{\|\mathbf{y}'\|\le\|\mathbf{x}\|/(2\overline{\mu}\sqrt{2t})} e^{-\|\mathbf{y}'\|^2}\, g(\|\mathbf{x}\|/2)\,d\mathbf{y}' + g(0)\int_{\|\mathbf{y}'\|>\|\mathbf{x}\|/(2\overline{\mu}\sqrt{2t})} e^{-\|\mathbf{y}'\|^2}\,d\mathbf{y}'\Big], \\ &\le \frac{\overline{\mu}^n}{\pi^{n/2}\underline{\mu}^n}\Big[\pi^{n/2} g(\|\mathbf{x}\|/2) + g(0)\int_{\|\mathbf{y}'\|>\|\mathbf{x}\|/(2\overline{\mu}\sqrt{2T})} e^{-\|\mathbf{y}'\|^2}\,d\mathbf{y}'\Big]. \end{aligned} \tag{51}$$

Call $h(\mathbf{x})$ the quantity given in the right-hand side of the last inequality. Since by (9) the function $g(\|\cdot\|)$ is in $L^\infty(\mathbb{R}^n) \cap L^1(\mathbb{R}^n)$, the function h belongs to $L^1(\mathbb{R}^n)$, and is independent of $\mathbf{z} \in \mathbb{R}_+^n$ and $t \in (0,T]$. One then infers from (50) and Lebesgue's dominated convergence theorem that the function $M_{\mathbf{p}}$ is continuous in $\mathbb{R}_+ \times \mathbb{R}_+^n$. As $C_{\mathbf{p}} = \log M_{\mathbf{p}}$, the cumulant generating function $C_{\mathbf{p}}$ is also continuous in $\mathbb{R}_+ \times \mathbb{R}_+^n$.

Let us then check that $M_{\mathbf{p}}$ and $C_{\mathbf{p}}$ are $C^{0,1}(\mathbb{R}_+ \times \mathbb{R}_+^n)$, meaning that the functions $\partial_i M_{\mathbf{p}} = \frac{\partial M_{\mathbf{p}}}{\partial z_i}$ and $\partial_i C_{\mathbf{p}} = \frac{\partial C_{\mathbf{p}}}{\partial z_i}$ exist and are continuous in $\mathbb{R}_+ \times \mathbb{R}_+^n$, for every $1 \le i \le n$. As a matter of fact, since $\mathbf{p}(t,\cdot)$ is a probability density function in $\mathbb{R}_-^n$ for any $t \ge 0$ and since the integral $\int_{\mathbb{R}_-^n} \|\mathbf{m}\|\,\mathbf{p}(t,\mathbf{m})\,d\mathbf{m}$ converges by formula (15) in Proposition 2.3, it easily follows from

Lebesgue's dominated convergence theorem that $\partial_i M_{\mathbf{p}}(t,\mathbf{z})$ exists for all $(t,\mathbf{z}) \in \mathbb{R}_+ \times \mathbb{R}_+^n$ and $1 \le i \le n$, with:

$$\partial_i M_{\mathbf{p}}(t,\mathbf{z}) = \int_{\mathbb{R}_-^n} m_i \, e^{\mathbf{z}\cdot\mathbf{m}} \mathbf{p}(t,\mathbf{m}) \, d\mathbf{m}, \tag{52}$$

hence, as in (49)-(50),

$$\begin{aligned} \partial_i M_{\mathbf{p}}(t,\mathbf{z}) &= \frac{2^{n/2}}{1+\int_0^t \overline{m}_v(s)ds} \int_{\mathbb{R}_-^n} m_i \frac{e^{\mathbf{z}\cdot\mathbf{m}}}{\sqrt{|m_1 \cdots m_n|}} v(t,\mathbf{x}^{\mathbf{1}}(\mathbf{m})) \, d\mathbf{m}, \\ &= -\frac{1}{1+\int_0^t \overline{m}_v(s)ds} \int_{\mathbb{R}^n} \frac{x_i^2}{2} \exp\Big(-\sum_{i=1}^n \frac{z_i x_i^2}{2}\Big) v(t,\mathbf{x}) \, d\mathbf{x}, \end{aligned} \tag{53}$$

from (16) and (48). On the one hand, the function $t \mapsto \int_0^t \overline{m}_v(s)ds$ is continuous in $\mathbb{R}_+$ and so is the function $(t,\mathbf{z}) \mapsto x_i^2 \exp(-\sum_{i=1}^n z_i x_i^2/2) \, v(t,\mathbf{x})$ in $\mathbb{R}_+ \times \mathbb{R}_+^n$, for every $\mathbf{x} \in \mathbb{R}^n$. On the other hand, as in the previous paragraph, it follows from (39)-(40) and (44) that, for any $1 \le i \le n$, $\mathbf{z} \in \mathbb{R}_+^n$, $T > 0$, $t \in (0,T]$ and $\mathbf{x} \in \mathbb{R}^n$, there holds:

$$\begin{aligned} &0 \le x_i^2 \exp\Big(-\sum_{i=1}^n \frac{z_i x_i^2}{2}\Big) v(t,\mathbf{x}), \\ &\le \|\mathbf{x}\|^2 v(t,\mathbf{x}) \le \frac{\overline{\mu}^n}{\pi^{n/2}\underline{\mu}^n}\Big[\pi^{n/2}\|\mathbf{x}\|^2 g(\|\mathbf{x}\|/2) + g(0)\|\mathbf{x}\|^2 \int_{\|\mathbf{y}'\|>\|\mathbf{x}\|/(2\overline{\mu}\sqrt{2T})} e^{-\|\mathbf{y}'\|^2} \, d\mathbf{y}'\Big]. \end{aligned}$$

Call $\widetilde{h}(\mathbf{x})$ the quantity given in the right-hand side of the last inequality. Since by (9) the function $\mathbf{x} \mapsto \|\mathbf{x}\|^2 g(\|\mathbf{x}\|)$ is in $L^1(\mathbb{R}^n)$, the function $\widetilde{h}$ belongs to $L^1(\mathbb{R}^n)$, and is independent of $\mathbf{z} \in \mathbb{R}_+^n$ and $t \in (0,T]$. One then infers from (53) and Lebesgue's dominated convergence theorem that the function $\partial_i M_{\mathbf{p}}$ is continuous in $\mathbb{R}_+ \times \mathbb{R}_+^n$, and so is the function $\partial_i C_{\mathbf{p}} = \partial_i M_{\mathbf{p}}/M_{\mathbf{p}}$.

In this paragraph, we are interested in the differentiation of $M_{\mathbf{p}}$ with respect to t. By (37), we already know that:

$$|\partial_t v(t,\mathbf{x})| \le \sum_{i=1}^n \frac{\mu_i^2}{2} |\partial_{ii} v(t,\mathbf{x})| + |m(\mathbf{x}) \, v(t,\mathbf{x})|, \tag{54}$$

for all $(t,\mathbf{x}) \in \mathbb{R}_+ \times \mathbb{R}^n$. Fix $T > 0$ and let $S > 0$ be the constant given in (38). Thus, for all $(t,\mathbf{x}) \in [0,T] \times \mathbb{R}^n$, there holds:

$$|\partial_{ii} v(t,\mathbf{x})| \le \|v(t,\cdot)\|_{C^{2+\alpha}(\mathbb{R}^n)} \le S\|q_0^{\#}\|_{C^{2+\alpha}(\mathbb{R}^n)} \le S\|q_0\|_{C^{2+\alpha}(\mathbb{R}^n)}. \tag{55}$$

Let us now focus on the boundedness of the second term of the right-hand side of (54), that is, the boundedness of the function $(t,\mathbf{x}) \mapsto m(\mathbf{x})v(t,\mathbf{x})$ in $[0,T] \times \mathbb{R}^n$. Since this function is continuous in $\mathbb{R}_+ \times \mathbb{R}^n$, let us show its boundedness in $(0,T] \times \mathbb{R}^n$. Thanks to (9) and (39)-(40), we get, as in (51), that:

$$|m(\mathbf{x})v(t,\mathbf{x})| \le \int_{\mathbb{R}^n} \frac{\overline{\mu}^n \|x\|^2}{2\pi^{n/2}\underline{\mu}^n} q_0^{\#}\big(\mathbf{x} - \overline{\mu}\sqrt{2t}\mathbf{y}\big) e^{-\|\mathbf{y}\|^2} \, d\mathbf{y},$$

for all $(t,\mathbf{x}) \in (0,T] \times \mathbb{R}^n$. Thus, we have:

$$\begin{aligned} |m(\mathbf{x})v(t,\mathbf{x})| &\le \frac{\overline{\mu}^n}{\pi^{n/2}\underline{\mu}^n} \int_{\mathbb{R}^n} \big(\|\mathbf{x} - \overline{\mu}\sqrt{2t}\mathbf{y}\|^2 + 2t\overline{\mu}^2\|\mathbf{y}\|^2\big) g\big(\|\mathbf{x} - \overline{\mu}\sqrt{2t}\mathbf{y}\|\big) e^{-\|\mathbf{y}\|^2} \, d\mathbf{y}, \\ &\le \frac{C\,\overline{\mu}^n}{\pi^{n/2}\underline{\mu}^n} + \frac{2\overline{\mu}^{n+2} T \, g(0)}{\pi^{n/2}\underline{\mu}^n} \int_{\mathbb{R}^n} \|\mathbf{y}\|^2 e^{-\|\mathbf{y}\|^2} \, d\mathbf{y}, \end{aligned}$$

where the constant C is such that $\|\mathbf{x}'\|^2 q_0^{\#}(\mathbf{x}') \le \|\mathbf{x}'\|^2 g(\|\mathbf{x}'\|) \le C$, for all $\mathbf{x}' \in \mathbb{R}^n$. Therefore, the function $(t,\mathbf{x}) \mapsto m(\mathbf{x})v(t,\mathbf{x})$ is bounded in $[0,T]\times\mathbb{R}^n$ for any $T>0$, and so is $\partial_t v$ by (54)-(55). Together with (38), (48) and the continuity of $\overline{m}_v$ in $\mathbb{R}_+$, it follows that the function $\partial_t q^{\#}$ is bounded in $[0,T]\times\mathbb{R}^n$, for every $T>0$. Finally, (16) implies that for all $(t,\mathbf{z}) \in [0,T]\times\mathbb{R}^n_+$ and $\mathbf{m} \in (\mathbb{R}^*_-)^n$,

$$|e^{\mathbf{z}\cdot\mathbf{m}}\ \partial_t \mathbf{p}(t,\mathbf{m})| \le 2^{n/2}\|\partial_t q^{\#}\|_{L^\infty([0,T]\times\mathbb{R}^n)} \frac{e^{\mathbf{z}\cdot\mathbf{m}}}{\sqrt{|m_1\cdots m_n|}}.$$

Since the integrals:

$$\int_{\mathbb{R}^n_-} \frac{e^{\mathbf{z}\cdot\mathbf{m}}}{\sqrt{|m_1\cdots m_n|}}\, d\mathbf{m} = 2^{n/2}\int_{\mathbb{R}^n_+} \exp\Big(-\sum_{i=1}^n \frac{z_i x_i^2}{2}\Big)\, d\mathbf{x},$$

converge for all $\mathbf{z} \in (\mathbb{R}^*_+)^n$, it then easily follows from the previous estimates and from Lebesgue's dominated convergence theorem that the function $M_{\mathbf{p}}$ is differentiable with respect to t in $\mathbb{R}_+\times(\mathbb{R}^*_+)^n$, with:

$$\partial_t M_{\mathbf{p}}(t,\mathbf{z}) = \int_{\mathbb{R}^n_-} e^{\mathbf{z}\cdot\mathbf{m}}\ \partial_t \mathbf{p}(t,\mathbf{m})\, d\mathbf{m}, \tag{56}$$

and that the function $\partial_t M_{\mathbf{p}}$ is itself continuous in $\mathbb{R}_+\times(\mathbb{R}^*_+)^n$. So is the function $\partial_t C_{\mathbf{p}} = \partial_t M_{\mathbf{p}}/M_{\mathbf{p}}$. The continuity of the functions $\partial_t M_{\mathbf{p}}$ and $\partial_t C_{\mathbf{p}}$ in the closure $\mathbb{R}_+\times\mathbb{R}^n_+$ of $\mathbb{R}_+\times(\mathbb{R}^*_+)^n$ will be obtained as a consequence of the equations satisfied by these two functions, which shall be established below.

Let us then turn to find an equation satisfied by $M_{\mathbf{p}}$, in order to derive the equation (23) satisfied by $C_{\mathbf{p}}$. Fix $(t,\mathbf{z}) \in \mathbb{R}_+\times(\mathbb{R}^*_+)^n$. Thanks to (17), we have:

$$\begin{aligned} e^{\mathbf{z}\cdot\mathbf{m}}\,\partial_t \mathbf{p}(t,\mathbf{m}) = &-\sum_{i=1}^n \mu_i^2 m_i e^{\mathbf{z}\cdot\mathbf{m}}\partial_{ii}\mathbf{p}(t,\mathbf{m}) - \frac{3}{2}\sum_{i=1}^n \mu_i^2 e^{\mathbf{z}\cdot\mathbf{m}}\partial_i \mathbf{p}(t,\mathbf{m}) \\ &+ \Big(\sum_{i=1}^n m_i - \overline{m}(t)\Big) e^{\mathbf{z}\cdot\mathbf{m}}\ \mathbf{p}(t,\mathbf{m}), \end{aligned} \tag{57}$$

for all $\mathbf{m} \in (\mathbb{R}^*_-)^n$.

We are now going to integrate (57) over $\mathbb{R}^n_-$. To do so, let us first focus on the first two terms of the right-hand side of (57). Fix an index $i \in [\![1,n]\!]$ and consider the cubes:

$$B_\varepsilon = \big\{\mathbf{m} = (m_1,\dots,m_n) \in \mathbb{R}^n_-,\ -\varepsilon^{-1} < m_j < -\varepsilon,\ \text{for each } j \in [\![1,n]\!]\big\} = (-\varepsilon^{-1},-\varepsilon)^n,$$

with $0<\varepsilon<1$. Denote $\widehat{B}_\epsilon = (-\varepsilon^{-1},-\varepsilon)^{n-1}$ and:

$$\widehat{\mathbf{m}} = (m_1,\dots,m_{i-1},m_{i+1},\dots,m_n), \quad \mathbf{m}^\rho = (m_1,\dots,m_{i-1},-\rho,m_{i+1},\dots,m_n),$$

for $\rho \in \mathbb{R}$. By using Fubini's theorem and integrating by parts with respect to the variable m_i, one infers that:

$$\begin{aligned} &\int_{B_\varepsilon}\Big(m_i e^{\mathbf{z}\cdot\mathbf{m}}\partial_{ii}\mathbf{p}(t,\mathbf{m}) + \frac{3}{2}e^{\mathbf{z}\cdot\mathbf{m}}\partial_i\mathbf{p}(t,\mathbf{m})\Big)\, d\mathbf{m} \\ &= \int_{B_\varepsilon}\Big(\frac{z_i}{2} + m_i z_i^2\Big) e^{\mathbf{z}\cdot\mathbf{m}}\,\mathbf{p}(t,\mathbf{m})\, d\mathbf{m} \\ &+ \int_{\widehat{B}_\varepsilon}\Big[z_i\,\varepsilon\, e^{\mathbf{z}\cdot\mathbf{m}^\varepsilon}\mathbf{p}(t,\mathbf{m}^\varepsilon) - z_i\,\varepsilon^{-1}\, e^{\mathbf{z}\cdot\mathbf{m}^{1/\varepsilon}}\mathbf{p}(t,\mathbf{m}^{1/\varepsilon})\Big]\, d\widehat{\mathbf{m}} \\ &+ \int_{\widehat{B}_\varepsilon}\Big[e^{\mathbf{z}\cdot\mathbf{m}^\varepsilon}\Big(\frac{\mathbf{p}(t,\mathbf{m}^\varepsilon)}{2} - \varepsilon\partial_i\mathbf{p}(t,\mathbf{m}^\varepsilon)\Big) - e^{\mathbf{z}\cdot\mathbf{m}^{1/\varepsilon}}\Big(\frac{\mathbf{p}(t,\mathbf{m}^{1/\varepsilon})}{2} - \varepsilon^{-1}\partial_i\mathbf{p}(t,\mathbf{m}^{1/\epsilon})\Big)\Big]\, d\widehat{\mathbf{m}}. \end{aligned} \tag{58}$$

Let us pass to the limit as $\varepsilon \to 0$ in the three integrals of the right-hand side of (58). Firstly, since $\mathbf{p}(t,\cdot)$ is nonnegative and the functions $\mathbf{m} \mapsto \mathbf{p}(t,\mathbf{m})$ and $\mathbf{m} \mapsto m_i \mathbf{p}(t,\mathbf{m})$ are in $L^1(\mathbb{R}^n_-)$, it follows from Lebesgue's dominated convergence theorem together with (21) and (52) that:

$$\int_{B_\varepsilon} \Big(\frac{z_i}{2} + m_i z_i^2 \Big) e^{\mathbf{z}\cdot\mathbf{m}} \, \mathbf{p}(t,\mathbf{m}) \, d\mathbf{m} \to \int_{\mathbb{R}^n_-} \Big(\frac{z_i}{2} + m_i z_i^2 \Big) e^{\mathbf{z}\cdot\mathbf{m}} \, \mathbf{p}(t,\mathbf{m}) \, d\mathbf{m} = \frac{z_i}{2} M_{\mathbf{p}}(t,\mathbf{z}) + z_i^2 \partial_i M_{\mathbf{p}}(t,\mathbf{z}),$$

as $\varepsilon \to 0$. Secondly, by denoting:

$$\widehat{C}_\varepsilon = \big(\sqrt{2\varepsilon}, \sqrt{2\varepsilon^{-1}} \big)^{n-1}, \quad \widehat{\mathbf{z}} = (z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_n),$$

and:

$$\widehat{\mathbf{x}} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n), \quad \mathbf{x}^\rho = (x_1, \ldots, x_{i-1}, \sqrt{2\rho}, x_{i+1}, \ldots, x_n),$$

for $\rho \geq 0$, it follows from (16) that:

$$\begin{aligned} \int_{\widehat{B}_\varepsilon} z_i \, \varepsilon \, e^{\mathbf{z}\cdot\mathbf{m}^\varepsilon} \mathbf{p}(t,\mathbf{m}^\varepsilon) \, d\widehat{\mathbf{m}} &= z_i \sqrt{\varepsilon} \, e^{-z_i \varepsilon} \int_{\widehat{B}_\varepsilon} \frac{2^{n/2} \, e^{\widehat{\mathbf{z}}\cdot\widehat{\mathbf{m}}} \, q^{\#}(t, \mathbf{x}^{\mathbb{1}}(\mathbf{m}^\varepsilon))}{\sqrt{|m_1 \cdots m_{i-1} m_{i+1} \cdots m_n|}} \, d\widehat{\mathbf{m}}, \\ &= z_i \sqrt{\varepsilon} \, e^{-z_i \varepsilon} \, 2^{n-1/2} \int_{\widehat{C}_\varepsilon} \exp\Big(- \sum_{j \neq i} \frac{z_j x_j^2}{2} \Big) q^{\#}(t, \mathbf{x}^\varepsilon) \, d\widehat{\mathbf{x}}. \end{aligned}$$

Since the continuous function $q^{\#}(t,\cdot)$ is bounded in $\mathbb{R}^n$ by (38) and (48), and since $\mathbf{z} \in (\mathbb{R}^*_+)^n$, one then gets that:

$$\int_{\widehat{B}_\varepsilon} z_i \, \varepsilon \, e^{\mathbf{z}\cdot\mathbf{m}^\varepsilon} \mathbf{p}(t,\mathbf{m}^\varepsilon) \, d\widehat{\mathbf{m}} \to 0, \quad \text{as } \varepsilon \to 0.$$

Similarly, we prove that:

$$\int_{\widehat{B}_\varepsilon} z_i \, \varepsilon^{-1} \, e^{\mathbf{z}\cdot\mathbf{m}^{1/\varepsilon}} \mathbf{p}(t,\mathbf{m}^{1/\varepsilon}) \, d\widehat{\mathbf{m}} = z_i \sqrt{\varepsilon^{-1}} \, e^{-z_i/\varepsilon} \, 2^{n-1/2} \int_{\widehat{C}_\varepsilon} \exp\Big(- \sum_{j \neq i} \frac{z_j x_j^2}{2} \Big) q^{\#}(t, \mathbf{x}^{1/\varepsilon}) \, d\widehat{\mathbf{x}} \to 0,$$

as $\varepsilon \to 0$. Thirdly, from the computations done in the proof of Theorem 2.4, we already know that, for all $\mathbf{m} \in (\mathbb{R}^*_-)^n$:

$$\frac{1}{2} \mathbf{p}(t,\mathbf{m}) + m_i \partial_i \mathbf{p}(t,\mathbf{m}) = \frac{2^{n/2} \partial_i q^{\#}(t, \mathbf{x}^{\mathbb{1}}(\mathbf{m}))}{\sqrt{2|m_1 \cdots m_{i-1} m_{i+1} \cdots m_n|}}.$$

Hence, we have:

$$\begin{aligned} \int_{\widehat{B}_\varepsilon} e^{\mathbf{z}\cdot\mathbf{m}^\varepsilon} \left(\frac{\mathbf{p}(t,\mathbf{m}^\varepsilon)}{2} - \varepsilon \partial_i \mathbf{p}(t,\mathbf{m}^\varepsilon) \right) d\widehat{\mathbf{m}} &= e^{-z_i \varepsilon} \int_{\widehat{B}_\varepsilon} \frac{2^{n/2-1/2} \, e^{\widehat{\mathbf{z}}\cdot\widehat{\mathbf{m}}} \, \partial_i q^{\#}(t, \mathbf{x}^{\mathbb{1}}(\mathbf{m}^\varepsilon))}{\sqrt{|m_1 \cdots m_{i-1} m_{i+1} \cdots m_n|}} \, d\widehat{\mathbf{m}}, \\ &= e^{-z_i \varepsilon} \, 2^{n-1} \int_{\widehat{C}_\varepsilon} \exp\left(- \sum_{j \neq i} \frac{z_j x_j^2}{2} \right) \partial_i q^{\#}(t, \mathbf{x}^\varepsilon) \, d\widehat{\mathbf{x}}. \end{aligned}$$

Since the function $\partial_i q^{\#}(t,\cdot)$ is continuous and bounded in $\mathbb{R}^n$ by (38) and (48), since $\partial_i q^{\#}(t,\mathbf{x}^0) = 0$, by #-symmetry of $q^{\#}(t,\cdot)$, and since $\mathbf{z} \in (\mathbb{R}^*_+)^n$, one then infers from Lebesgue's dominated convergence theorem that:

$$\int_{\widehat{B}_\varepsilon} e^{\mathbf{z}\cdot\mathbf{m}^\varepsilon} \left(\frac{\mathbf{p}(t,\mathbf{m}^\varepsilon)}{2} - \varepsilon \partial_i \mathbf{p}(t,\mathbf{m}^\varepsilon) \right) d\widehat{\mathbf{m}} \to 0, \quad \text{as } \varepsilon \to 0.$$

Furthermore, the integral:

$$\int_{\widehat{B}_\varepsilon} e^{\mathbf{z}\cdot\mathbf{m}^{1/\varepsilon}} \left(\frac{\mathbf{p}(t,\mathbf{m}^{1/\varepsilon})}{2} - \varepsilon^{-1}\partial_i \mathbf{p}(t,\mathbf{m}^{1/\varepsilon}) \right) d\widehat{\mathbf{m}} = e^{-z_i/\varepsilon}\, 2^{n-1} \int_{\widehat{C}_\varepsilon} \exp\left(-\sum_{j\neq i} \frac{z_j x_j^2}{2} \right) \partial_i q^{\#}(t,\mathbf{x}^{1/\varepsilon})\, d\widehat{\mathbf{x}},$$

converges to 0 as $\varepsilon \to 0$. Coming back to (58) and passing to the limit as $\varepsilon \to 0$, it follows from the previous estimates that:

$$\int_{B_\varepsilon} \left(m_i e^{\mathbf{z}\cdot\mathbf{m}} \partial_{ii}\mathbf{p}(t,\mathbf{m}) + \frac{3}{2} e^{\mathbf{z}\cdot\mathbf{m}} \partial_i \mathbf{p}(t,\mathbf{m}) \right) d\mathbf{m} \underset{\varepsilon\to 0}{\longrightarrow} \frac{z_i}{2} M_{\mathbf{p}}(t,\mathbf{z}) + z_i^2 \partial_i M_{\mathbf{p}}(t,\mathbf{z}). \tag{59}$$

Let us finally remember (57) and that the functions:

$$\mathbf{m} \mapsto e^{\mathbf{z}\cdot\mathbf{m}} \partial_t \mathbf{p}(t,\mathbf{m}), \quad \mathbf{m} \mapsto m_i e^{\mathbf{z}\cdot\mathbf{m}} \mathbf{p}(t,\mathbf{m}) \quad \text{and} \quad \mathbf{m} \mapsto e^{\mathbf{z}\cdot\mathbf{m}} \mathbf{p}(t,\mathbf{m}),$$

are in $L^1(\mathbb{R}^n_-)$ with integrals given by (56), (52) and (21), respectively. Together with (58) and (59), one concludes that:

$$\partial_t M_{\mathbf{p}}(t,\mathbf{z}) = \sum_{i=1}^n (1-\mu_i^2 z_i^2)\partial_i M_{\mathbf{p}}(t,\mathbf{z}) - \frac{1}{2}\sum_{i=1}^n \mu_i^2\, z_i\, M_{\mathbf{p}}(t,\mathbf{z}) - \overline{m}(t) M_{\mathbf{p}}(t,\mathbf{z}), \tag{60}$$

for every $(t,\mathbf{z}) \in \mathbb{R}_+ \times (\mathbb{R}_+^*)^n$. Since the right-hand side of the above equation is continuous in $\mathbb{R}_+ \times \mathbb{R}_+^n$, one infers that the function $\partial_t M_{\mathbf{p}}$ is extendable by continuity in $\mathbb{R}_+ \times \mathbb{R}_+^n$ and (60) holds in $\mathbb{R}_+ \times \mathbb{R}_+^n$. Owing to the definition $C_{\mathbf{p}} = \log M_{\mathbf{p}}$, one concludes that $\partial_t C_{\mathbf{p}}$ is continuous in $\mathbb{R}_+ \times \mathbb{R}_+^n$ (finally, $C_{\mathbf{p}}$ is of class $C^{1,1}(\mathbb{R}_+ \times \mathbb{R}_+^n)$) and:

$$\partial_t C_{\mathbf{p}}(t,\mathbf{z}) = \sum_{i=1}^n (1-\mu_i^2 z_i^2)\partial_i C_{\mathbf{p}}(t,\mathbf{z}) - \frac{1}{2}\sum_{i=1}^n \mu_i^2\, z_i - \overline{m}(t) = A(\mathbf{z})\cdot\nabla C_{\mathbf{p}}(t,\mathbf{z}) - b(\mathbf{z}) - \overline{m}(t),$$

for all $(t,\mathbf{z}) \in \mathbb{R}_+ \times \mathbb{R}_+^n$, where A and b are as in (24). Therefore, (23) holds in $\mathbb{R}_+ \times \mathbb{R}_+^n$ and the proof of Theorem 2.6 is thereby complete. □

Before going into the proof of the remaining results, let us first observe that, in (23)-(24), $\overline{m}(t) = \mathbf{1}\cdot\nabla C_{\mathbf{p}}(t,\mathbf{O})$, with:

$$\mathbf{1} = (1,\ldots,1) \in \mathbb{R}^n.$$

It turns out that, if $A(\mathbf{z}) = \mathbf{1}$, then equation (23) can be solved explicitely by the method of characteristics, as the following lemma shows (this lemma is used later in the proof of Proposition 2.8 in the general case $A(\mathbf{z})$ given in (24)).

Lemma 4.3. The Cauchy problem:

$$\begin{cases} \partial_t Q(t,\mathbf{z}) &= \mathbf{1}\cdot(\nabla Q(t,\mathbf{z}) - \nabla Q(t,\mathbf{O})) - \widetilde{b}(\mathbf{z}),\ t\geq 0,\ \mathbf{z}\in\mathbb{R}_+^n, \\ Q(0,\mathbf{z}) &= Q_0(\mathbf{z}),\ \mathbf{z}\in\mathbb{R}_+^n, \\ Q(t,\mathbf{O}) &= 0,\ t\geq 0, \end{cases} \tag{61}$$

with $\widetilde{b} \in C^1(\mathbb{R}_+^n)$ and $Q_0 \in C^1(\mathbb{R}_+^n)$ such that $\widetilde{b}(\mathbf{O}) = Q_0(\mathbf{O}) = 0$, admits a unique $C^{1,1}(\mathbb{R}_+ \times \mathbb{R}_+^n)$ solution, which is given by the expression:

$$Q(t,\mathbf{z}) = \int_0^t \left(\widetilde{b}(s\mathbf{1}) - \widetilde{b}(\mathbf{z}+s\mathbf{1}) \right) ds + Q_0(\mathbf{z}+t\mathbf{1}) - Q_0(t\mathbf{1}). \tag{62}$$

Proof. First of all, it is immediate to check that the function Q given by (62) is a $C^{1,1}(\mathbb{R}_+ \times \mathbb{R}_+^n)$ solution of (61). Let now Q_1 and Q_2 be two $C^{1,1}(\mathbb{R}_+ \times \mathbb{R}_+^n)$ solutions of (61) and denote $Q = Q_1 - Q_2$. The function Q is of class $C^{1,1}(\mathbb{R}_+ \times \mathbb{R}_+^n)$ and obeys:

$$\partial_t Q(t,\mathbf{z}) = \mathbf{1} \cdot (\nabla Q(t,\mathbf{z}) - \nabla Q(t,\mathbf{O})),$$

for all $(t,\mathbf{z}) \in \mathbb{R}_+ \times \mathbb{R}_+^n$, together with $Q(0,\mathbf{z}) = 0$ for all $\mathbf{z} \in \mathbb{R}_+^n$ and $Q(t,\mathbf{O}) = 0$ for all $t \geq 0$. It remains to show that $Q = 0$ in $\mathbb{R}_+ \times \mathbb{R}_+^n$. Fix any $(t,\mathbf{z}) \in \mathbb{R}_+ \times \mathbb{R}_+^n$. If $t = 0$, then $Q(0,\mathbf{z}) = 0$, so let us assume that $t > 0$. Consider the $C^1([0,t])$ function R defined by $R(s) = Q(t-s, \mathbf{z}+s\mathbf{1}) - Q(t-s, s\mathbf{1})$ for $s \in [0,t]$ (which is well defined since $\mathbf{z} + s\mathbf{1} \in \mathbb{R}_+^n$). It follows from the equation satisfied by Q that, for all $s \in [0,t]$, there holds:

$$R'(s) = -\partial_t Q(t-s, \mathbf{z}+s\mathbf{1}) - \partial_t Q(t-s, s\mathbf{1}) + \mathbf{1} \cdot (\nabla Q(t-s, \mathbf{z}+s\mathbf{1}) - \nabla Q(t-s, s\mathbf{1})) = 0.$$

Hence, $Q(t,\mathbf{z}) = Q(t,\mathbf{z}) - Q(t,\mathbf{O}) = R(0) = R(t) = Q(0, \mathbf{z}+t\mathbf{1}) - Q(0, t\mathbf{1}) = 0$, which is the desired conclusion. □

Proof of Proposition 2.8. In order to derive a general formula for the $C^{1,1}(\mathbb{R}_+ \times \mathbb{R}_+^n)$ solution $C_{\mathbf{p}}$ of (23), we make a substitution of the spatial variable and use the previous special case described in Lemma 4.3. To do so, we set, for $t \geq 0$ and $z \in \mathbb{R}_+^n$,

$$Q(t,\mathbf{z}) = C_{\mathbf{p}}(t, \mathbf{y}(\mathbf{z})),$$

where $\mathbf{y}(\mathbf{z}) = (y_1(\mathbf{z}), \ldots, y_n(\mathbf{z}))$ and $y_i(\mathbf{z}) = \tanh(\mu_i z_i)/\mu_i$ for every $1 \leq i \leq n$. Notice that $\mathbf{y}(\mathbf{z}) \in \mathbb{R}_+^n$ for every $\mathbf{z} \in \mathbb{R}_+^n$. The function Q is of class $C^{1,1}(\mathbb{R}_+ \times \mathbb{R}_+^n)$ and:

$$\mathbf{1} \cdot \nabla Q(t,\mathbf{z}) = \sum_{i=1}^n \partial_i Q(t,\mathbf{z}) = \sum_{i=1}^n \big(1 - \tanh^2(\mu_i z_i)\big) \partial_i C_{\mathbf{p}}(t, \mathbf{y}(\mathbf{z})) = A(\mathbf{y}(\mathbf{z})) \cdot \nabla C_{\mathbf{p}}(t, \mathbf{y}(\mathbf{z})),$$

for all $(t,\mathbf{z}) \in \mathbb{R}_+ \times \mathbb{R}_+^n$, where A is given in (24). As $\overline{m}(t) = \mathbf{1} \cdot \nabla C_{\mathbf{p}}(t,\mathbf{O}) = \mathbf{1} \cdot \nabla Q(t,\mathbf{O})$ and $Q(t,\mathbf{O}) = C_{\mathbf{p}}(t,\mathbf{O}) = \log M_{\mathbf{p}}(t,\mathbf{O}) = 0$ by (15) and (21), it follows from (23) that:

$$\begin{cases} \partial_t Q(t,\mathbf{z}) &= \mathbf{1} \cdot (\nabla Q(t,\mathbf{z}) - \nabla Q(t,\mathbf{O})) - b(\mathbf{y}(\mathbf{z})), \ t \geq 0, \ \mathbf{z} \in \mathbb{R}_+^n, \\ Q(t,\mathbf{O}) &= 0, \ t \geq 0, \end{cases}$$

and $Q(0,\mathbf{z}) = C_{\mathbf{p}_0}(\mathbf{y}(\mathbf{z}))$ for all $\mathbf{z} \in \mathbb{R}_+^n$. The functions $C_{\mathbf{p}_0} \circ \mathbf{y}$ and $\widetilde{b} := b \circ \mathbf{y}$ are of class $C^1(\mathbb{R}_+^n)$ and $C_{\mathbf{p}_0}(\mathbf{y}(\mathbf{O})) = \widetilde{b}(\mathbf{O}) = 0$. Therefore, Lemma 4.3 implies that:

$$Q(t,\mathbf{z}) = \int_0^t [b(\mathbf{y}(s\mathbf{1})) - b(\mathbf{y}(\mathbf{z}+s\mathbf{1}))]\, ds + C_{\mathbf{p}_0}(\mathbf{y}(\mathbf{z}+t\mathbf{1})) - C_{\mathbf{p}_0}(\mathbf{y}(t\mathbf{1})), \tag{63}$$

for all $(t,\mathbf{z}) \in \mathbb{R}_+ \times \mathbb{R}_+^n$. Consider now any $t \in \mathbb{R}_+$ and $\mathbf{z} = (z_1, \ldots, z_n) \in [0, 1/\mu_1) \times \cdots \times [0, 1/\mu_n)$. Set:

$$\mathbf{z}' = \Big(\frac{\operatorname{atanh}(\mu_1 z_1)}{\mu_1}, \ldots, \frac{\operatorname{atanh}(\mu_1 z_1)}{\mu_1}\Big) \in \mathbb{R}_+^n,$$

and observe that $\mathbf{y}(\mathbf{z}') = \mathbf{z}$. Hence, we have:

$$\begin{aligned} C_{\mathbf{p}}(t,\mathbf{z}) &= C_{\mathbf{p}}(t, \mathbf{y}(\mathbf{z}')), \\ &= Q(t,\mathbf{z}') = \int_0^t \big[b(\mathbf{y}(s\mathbf{1})) - b(\mathbf{y}(\mathbf{z}'+s\mathbf{1}))\big]\, ds + C_{\mathbf{p}_0}(\mathbf{y}(\mathbf{z}'+t\mathbf{1})) - C_{\mathbf{p}_0}(\mathbf{y}(t\mathbf{1})), \end{aligned}$$

which leads straightforwardly to the formulae (26)-(27). Furthermore, for every $t \in \mathbb{R}_+$, the formula $\overline{m}(t) = \mathbf{1} \cdot \nabla Q(t,\mathbf{O})$ together with (63) easily yields (28). The proof of Proposition 2.8 is thereby complete. □

Proof of Corollary 2.7. Let p be the fitness distribution, that is, $p(t,m)\,dm$ is the pushforward measure of $q(t,\mathbf{x})\,d\mathbf{x}$ by the map $\mathbf{x} \mapsto -\|\mathbf{x}\|^2/2$. Let:

$$M_p(t,z) = \int_{-\infty}^{0} e^{zm}\, p(t,m)\, dm,$$

be the moment generating function of p. As the fitness $m \in \mathbb{R}_-$ is the sum of the fitness components $(m_1,\dots,m_n) \in \mathbb{R}_-^n$, we have:

$$M_p(t,z) = \int_{-\infty}^{0} e^{zm}\, p(t,m)\, dm = \int_{\mathbb{R}_-^n} e^{z(m_1+\cdots+m_n)}\, \mathbf{p}(t,\mathbf{m})\, d\mathbf{m} = M_{\mathbf{p}}(t,z\mathbf{1}),$$

for all $(t,z) \in \mathbb{R}_+ \times \mathbb{R}_+$. This implies that $C_p(t,z) = C_{\mathbf{p}}(t,z\mathbf{1})$ for all $(t,z) \in \mathbb{R}_+ \times \mathbb{R}_+$ and that C_p is of class $C^{1,1}(\mathbb{R}_+ \times \mathbb{R}_+)$, with initial condition $C_p(0,\cdot) = C_{\mathbf{p}_0}(\cdot\mathbf{1})$. Thanks to the equations (23)-(24) satisfied by $C_{\mathbf{p}}$, it follows that:

$$\begin{aligned}\partial_t C_p(t,z) &= \sum_{i=1}^{n}\left[(1-\mu^2 z^2)\partial_i C_{\mathbf{p}}(t,z\mathbf{1}) - \partial_i C_{\mathbf{p}}(t,\mathbf{O}) - \frac{\mu^2}{2}\, z\right],\\ &= (1-\mu^2 z^2)\partial_z C_p(t,z) - \partial_z C_p(t,0) - \frac{n}{2}\mu^2 z,\end{aligned}$$

for all $(t,z) \in \mathbb{R}_+ \times \mathbb{R}_+$. This is the desired result and the proof of Corollary 2.7 is thereby complete. □

Proof of Corollary 2.9. We have seen in the proof of Corollary 2.7 that $C_p(t,z) = C_{\mathbf{p}}(t,z\mathbf{1})$ for all $(t,z) \in \mathbb{R}_+ \times \mathbb{R}_+$. Thus, formulae (26)-(28) straightforwardly yield (29)-(30) for $t \geq 0$ and $z \in [0, 1/\mu)$. □

4.4 Stationary states

Proof of Theorem 2.10. We use the notations of Proposition 2.8. Let $\mathbf{z} \in [0,1/\mu_1) \times \cdots \times [0,1/\mu_n)$. As $\psi(t,\mathbf{z}) \to (\mu_1^{-1},\dots,\mu_n^{-1})$ as $t \to +\infty$, the continuity of $C_{\mathbf{p}_0}$ and the formulae (26)-(27) yield that:

$$C_{\mathbf{p}}(t,\mathbf{z}) \to \frac{1}{2}\sum_{i=1}^{n} \log\left[\exp\left(-\operatorname{atanh}(\mu_i z_i)\right)\cosh\left(\operatorname{atanh}(\mu_i z_i)\right)\right] = -\frac{1}{2}\sum_{i=1}^{n}\log(1+\mu_i z_i), \tag{64}$$

as $t \to +\infty$. It then follows from the generalization of the Curtiss theorem [46] that, if the limit as $t \to +\infty$ of the cumulant generating functions $C_{\mathbf{p}}(t,\cdot)$ is the cumulant generating function of $\mathbf{p}_\infty$ given by (31) in some subset of $\mathbb{R}_-^n$ with non-empty interior, then the distributions $\mathbf{p}(t,\cdot)$ weakly converge to $\mathbf{p}_\infty$ as $t \to +\infty$. So let us compute the CGF of $\mathbf{p}_\infty$. For all $\mathbf{z} \in \mathbb{R}_+^n$, Fubini's theorem yields that:

$$\begin{aligned}\int_{\mathbb{R}_-^n} e^{\mathbf{z}\cdot\mathbf{m}}\mathbf{p}_\infty(\mathbf{m})\, d\mathbf{m} &= \frac{1}{\pi^{n/2}}\left(\prod_{i=1}^{n}\mu_i^{-1/2}\right)\int_{\mathbb{R}_-^n}\left(\prod_{i=1}^{n}|m_i|^{-1/2}\right)\exp\left(\sum_{i=1}^{n} m_i/\mu_i\right)e^{\mathbf{z}\cdot\mathbf{m}}\, d\mathbf{m},\\ &= \frac{1}{\pi^{n/2}}\prod_{i=1}^{n}\left[\mu_i^{-1/2}\int_{-\infty}^{0}|m_i|^{-1/2}\exp\left((1+\mu_i z_i)\, m_i/\mu_i\right)\, dm_i\right],\\ &= \frac{1}{\pi^{n/2}}\prod_{i=1}^{n}\left[\frac{1}{\sqrt{1+\mu_i z_i}}\int_{0}^{+\infty}|x_i|^{-1/2}\, e^{-x_i}\, dx_i\right] = \prod_{i=1}^{n}(1+\mu_i z_i)^{-1/2}.\end{aligned}$$

Hence, the CGF of $\mathbf{p}_\infty$ is equal to the function $\mathbf{z} = (z_1, \dots, z_n) \mapsto -(1/2)\sum_{i=1}^n \log(1+\mu_i z_i)$, that is, the limit in (64). As a consequence, the distributions $\mathbf{p}(t,\cdot)$ weakly converge to $\mathbf{p}_\infty$ in $\mathbb{R}^n_-$ as $t \to +\infty$.

On the other hand, thanks to Proposition 2.8, we also know that:

$$\overline{m}(t) = A(\psi(t,\mathbf{O})) \cdot \nabla C_{\mathbf{p}_0}(\psi(t,\mathbf{O})) - b(\psi(t,\mathbf{O})),$$

for every $t \geq 0$, with $A(\mathbf{z}) = (1-\mu_1^2 z_1^2, \dots, 1-\mu_n^2 z_n^2)$ and $b(\mathbf{z}) = \sum_{i=1}^n \mu_i^2 z_i/2$. Notice that $\psi(t,\mathbf{O}) \to (\mu_1^{-1}, \dots, \mu_n^{-1})$, $A(\psi(t,\mathbf{O})) \to \mathbf{O}$ and $b(\psi(t,\mathbf{O})) \to \sum_{i=1}^n \mu_i/2$, as $t \to +\infty$. Hence, $\overline{m}(t) \to -\sum_{i=1}^n \mu_i/2 =: \overline{m}_\infty$, as $t \to +\infty$. It is also straightforward to check that:

$$\sum_{i=1}^n \int_{\mathbb{R}^n_-} m_i\, \mathbf{p}_\infty(\mathbf{m})\, d\mathbf{m} = \overline{m}_\infty,$$

and that $\mathbf{p}_\infty$ is a classical $C^2((\mathbb{R}^*_-)^n)$ solution of (32) (this property is also a consequence of the fact that $\mathbf{p}$ satisfies (17) and the distributions $\mathbf{p}(t,\cdot)$ weakly converge to $\mathbf{p}_\infty$ as $t \to +\infty$). The proof of Theorem 2.10 is thereby complete. □

Proof of Corollary 2.11. By the same arguments as in the proof of Theorem 2.10, thanks to (29), we see that, for all $z \in [0, 1/\mu)$, $C_p(t,z) \to -(1/2)\log(1+\mu z)$ as $t \to +\infty$. This limiting function corresponds to the cumulant generating function of a random variable distributed according to $-\Gamma(n/2,\mu)$. Since a distribution is uniquely determined by its cumulant generating function, this implies that p_∞ is the probability density function of this random variable, *i.e.*, for all $m < 0$,

$$p_\infty(m) = \frac{1}{\Gamma(n/2)\,\mu^{n/2}}\, |m|^{\frac{n}{2}-1} \exp\left(\frac{m}{\mu}\right),$$

where $\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt$ is the standard Gamma function. □

Proof of Corollary 2.12. We assume here that q_0 is #-symmetric, that is, $q_0 = q_0^{\#}$. As already emphasized in Section 2.1, the uniqueness for problem (10) implies that $q(t,\cdot)$ is also #-symmetric at each time $t \geq 0$. Proposition 2.3 (or formula (16)) yields:

$$q(t,\mathbf{x}) = 2^{-n}|x_1 \cdots x_n|\, p\Big(t, -\frac{x_1^2}{2}, \dots, -\frac{x_n^2}{2}\Big),$$

for all $(t,\mathbf{x}) \in \mathbb{R}_+ \times (\mathbb{R}^*)^n$ and, for each function $\phi \in C_c^\infty((\mathbb{R}^*)^n)$:

$$\begin{aligned}
\int_{\mathbb{R}^n} q(t,\mathbf{x})\,\phi(\mathbf{x})\,d\mathbf{x} &= 2^{-n}\int_{\mathbb{R}^n} |x_1 \cdots x_n|\, p\Big(t, -\frac{x_1^2}{2}, \dots, -\frac{x_n^2}{2}\Big)\,\phi(\mathbf{x})\,d\mathbf{x}, \\
&= \int_{\mathbb{R}^n_+} |x_1 \cdots x_n|\, p\Big(t, -\frac{x_1^2}{2}, \dots, -\frac{x_n^2}{2}\Big)\,\phi^{\#}(\mathbf{x})\,d\mathbf{x}, \\
&= \int_{\mathbb{R}^n_-} \mathbf{p}(t,\mathbf{m})\,\phi^{\#}(\mathbf{x}^{\mathbf{1}}(\mathbf{m}))\,d\mathbf{m}, \\
&\to \int_{\mathbb{R}^n_-} \mathbf{p}_\infty(\mathbf{m})\,\phi^{\#}(\mathbf{x}^{\mathbf{1}}(\mathbf{m}))\,d\mathbf{m} = \int_{\mathbb{R}^n} q_\infty(\mathbf{x})\,\phi(\mathbf{x})\,d\mathbf{x},
\end{aligned}$$

as $t \to +\infty$, with $\mathbf{x}^{\mathbf{1}}(\mathbf{m}) = (\sqrt{-2m_1}, \dots, \sqrt{-2m_n})$ and:

$$q_\infty(\mathbf{x}) = 2^{-n}\,|x_1 \cdots x_n|\,\mathbf{p}_\infty\Big(-\frac{x_1^2}{2}, \dots, -\frac{x_n^2}{2}\Big).$$

The above formula corresponds to (33). Furthermore, since $\int_{\mathbb{R}^n} q_\infty(\mathbf{x})\,d\mathbf{x} = 1$ and since $\int_{\mathbb{R}^n} q(t,\mathbf{x})\,d\mathbf{x} = 1$, for every $t \geq 0$, it then easily follows from the previous estimates that:

$$\int_{\mathbb{R}^n} q(t,\mathbf{x})\,\phi(\mathbf{x})\,d\mathbf{x} \underset{t\to+\infty}{\longrightarrow} \int_{\mathbb{R}^n} q_\infty(\mathbf{x})\,\phi(\mathbf{x})\,d\mathbf{x},$$

for every $\phi \in C_c^\infty(\mathbb{R}^n)$. In other words, the distributions $q(t,\cdot)$ weakly converge in $\mathbb{R}^n$ to q_∞ as $t \to +\infty$. Lastly, the formula $\lim_{t\to+\infty} \overline{m}(t) = \overline{m}_\infty = \int_{\mathbb{R}^n} m(\mathbf{x})\,q_\infty(\mathbf{x})\,d\mathbf{x}$ is a consequence of the previous arguments and Theorem 2.10. The proof of Corollary 2.12 is thereby complete. □

4.5 Plateaus: proofs of Proposition 2.13 and Remark 2.14

Proof of Proposition 2.13. We show in this section that, given an initial phenotype $\mathbf{x}_0 = (x_{0,1},\ldots,x_{0,n}) \in \mathbb{R}^n$, a value $\mu_1 > 0$, and a duration $T > 0$, we can choose some positive real numbers $\mu_2, \ldots, \mu_n$ (here, $n \geq 2$) such that the mean fitness:

$$t \mapsto \overline{m}(t) := \sum_{i=1}^n \Big(\frac{x_{0,i}^2}{2}\big(\tanh^2(\mu_i t) - 1\big) - \frac{\mu_i}{2}\tanh(\mu_i t)\Big),$$

is close to each of the plateaus:

$$\overline{m}_{\infty,k} = -\sum_{i=k+1}^n \frac{x_{0,i}^2}{2} - \sum_{i=1}^k \frac{\mu_i}{2}, \quad \text{for } k = 1,\ldots,n-1,$$

at least during a time interval of duration T. We recall that the above formula for $\overline{m}(t)$ corresponds to the limit of formula (28), when the initial conditions q_0 approach the Dirac distribution $\delta_{\mathbf{x}_0}$.

More precisely, we are given $\mathbf{x}_0 = (x_{0,1},\ldots,x_{0,n}) \in \mathbb{R}^n$, $T > 0$, $\varepsilon > 0$ and $\mu_1 > 0$, and we shall prove the existence of $(\mu_2,\ldots,\mu_n) \in (\mathbb{R}_+^*)^{n-1}$ such that, for each $k \in [\![1, n-1]\!]$, the set:

$$\big\{t \geq 0,\ |\overline{m}(t) - \overline{m}_{\infty,k}| \leq \varepsilon\big\}, \tag{65}$$

contains an interval of length at least equal to T. In that respect, we firstly define some functions:

$$s_i : (\mu,t) \in \mathbb{R}_+^* \times \mathbb{R}_+ \mapsto \Big|\frac{x_{0,i}^2}{2}\tanh(\mu\,t) - \frac{\mu}{2}\Big| \quad \text{for } i = 1,\ldots,n-1.$$

Secondly, by iteration for $k = 1,\ldots,n-1$, we can then define:

- ◇ a function $S_k : t \in \mathbb{R}_+ \mapsto \sum_{i=1}^k \Big|\frac{x_{0,i}^2}{2}\big(\tanh(\mu_i t) + 1\big) - \frac{\mu_i}{2}\Big|$,
- ◇ a time $\tau_k > \tau_{k-1} + T$, (with $\tau_0 = -T$) such that:

$$\Big(\max_{t\in[\tau_k,\tau_k+T]} S_k(t)\Big) \times (1 - \tanh(\mu_k\tau_k)) \leq \varepsilon\frac{n+k}{n}, \tag{66}$$

- ◇ a real number $\mu_{k+1} \in (0,\mu_k)$ such that:

$$\forall\, i \in [\![1,k]\!],\ \forall\, t \in [\tau_i, \tau_i + T], \quad s_{k+1}(\mu_{k+1},t)\,\tanh(\mu_{k+1}(\tau_i + T)) \leq \frac{\varepsilon}{2n}.$$

Note that the last property implies that:

$$\forall\, k \in [\![1, n-1]\!],\ \forall\, i \in [\![k+1, n]\!],\ \forall\, t \in [\tau_k, \tau_k + T],\ s_i(\mu_i, t)\, \tanh(\mu_i(\tau_k + T)) \le \frac{\varepsilon}{2n}. \tag{67}$$

Fix now an index $k \in [\![1, n-1]\!]$ and a time $t \in [\tau_k, \tau_k + T]\,(\subset \mathbb{R}_+^*)$. There holds:

$$\begin{aligned} |\overline{m}(t) - \overline{m}_{\infty,k}| &= \Big| \sum_{i=1}^{k} \Big(\frac{x_{0,i}^2}{2} \big(\tanh^2(\mu_i t) - 1 \big) - \frac{\mu_i}{2} \big(\tanh(\mu_i t) - 1 \big) \Big) \\ &\qquad + \sum_{i=k+1}^{n} \Big(\frac{x_{0,i}^2}{2} \tanh^2(\mu_i t) - \frac{\mu_i}{2} \tanh(\mu_i t) \Big) \Big|, \\ &\le \sum_{i=1}^{k} \Big| \frac{x_{0,i}^2}{2} \big(\tanh(\mu_i t) + 1 \big) - \frac{\mu_i}{2} \Big| \big(1 - \tanh(\mu_i t) \big) \\ &\qquad + \sum_{i=k+1}^{n} \Big| \frac{x_{0,i}^2}{2} \tanh(\mu_i t) - \frac{\mu_i}{2} \Big| \tanh(\mu_i t). \end{aligned} \tag{68}$$

As $t \in [\tau_k, \tau_k + T]$, we have $0 < 1 - \tanh(\mu_i t) \le 1 - \tanh(\mu_k \tau_k)$ for every $i \in [\![1, k]\!]$ (remember that $0 < \mu_k < \mu_{k-1} < \cdots < \mu_1$), whereas $\tanh(\mu_i\, t) \le \tanh(\mu_i\, (\tau_k + T))$ for every $i \in [\![k+1, n]\!]$. It then follows from (66)-(68) that, for every $k \in [\![1, n-1]\!]$ and $t \in [\tau_k, \tau_k + T]$:

$$|\overline{m}(t) - \overline{m}_{\infty,k}| \le S_k(t) \big(1 - \tanh(\mu_k \tau_k) \big) + \sum_{i=k+1}^{n} s_i(\mu_i, t)\, \tanh(\mu_i\, (\tau_k + T)) \le \varepsilon \frac{n+k}{2n} + \sum_{i=k+1}^{n} \frac{\varepsilon}{2n} = \varepsilon.$$

Thus, with this choice of $(\mu_1, \ldots, \mu_n) \in (\mathbb{R}_+^*)^n$, each set defined in (65) contains an interval of length at least equal to T. This proves Proposition 2.13. □

Proof of results in Remark 2.14. Take $n = 2$, $\mu_1 = 1$, $\mu_2 = 10^{-k}$ for some $k \ge 1$ and $x_{0,1} = x_{0,2} = 1$. Differentiating the expression 34, we observe that:

$$\overline{m}'(t) = \left(\tanh(t) - \frac{1}{2} \right) (1 - \tanh^2(t)) + \mu_2 \left(\tanh(\mu_2\, t) - \frac{\mu_2}{2} \right) (1 - \tanh^2(\mu_2\, t)).$$

We observe that, as $t \to 0$, $\overline{m}'(t) \to -1/2 - 10^{-2\,k}/2$. Then $\overline{m}'(t)$ rapidly increases, to reach significantly positive values, *e.g.*, at $t_0 = \ln(3)$, $\overline{m}'(t_0) > 27/250 - \mu_2/2 > 0.1 - 10^{-k}/2$.

Then, consider the interval:

$$I := \left(\ln\left(2 \cdot 10^k\right), 10^{k/2} \right).$$

For each t in I, $1 - \tanh^2(t) < 4\, \exp(-2\, t) < 10^{-2\,k}$ and $\tanh(\mu_2\, t) - \mu_2/2 < \mu_2\, t < 10^{-k/2}$. Thus,

$$|\overline{m}'(t)| \le 10^{-2\,k} + 10^{-3\,k/2},$$

which means that $\overline{m}(t)$ remains stable within this interval, corresponding to a part of the plateau. Latter on, $\overline{m}'(t)$ reaches again significantly positive values which are significantly higher than in this interval, *e.g.*, at $t_1 = 10^k\, \ln[(\sqrt{2} + \sqrt{6})/2] > 10^{k/2}$, straightforward computations show that:

$$\begin{aligned} \overline{m}'(t_1) &\ge \mu_2 \left(\tanh(\mu_2\, t_1) - \tfrac{\mu_2}{2} \right) (1 - \tanh^2(\mu_2\, t_1)) \\ &= 2 \cdot 10^{-k} \left[\tfrac{(2+\sqrt{3})(2+2\sqrt{3})}{(3+\sqrt{3})^3} - 10^{-k} \tfrac{(2+\sqrt{3})}{(3+\sqrt{3})^2} \right], \end{aligned}$$

which show that $\overline{m}'(t)$ is of order 10^{-k} at t_1, vs $10^{-3k/2}$ in the interval I. □

A A formal derivation of the diffusive approximation of the mutation effects

The goal of this appendix is to give a formal justification of the diffusion term in (4). The case $n = 1$ is classical and can be found *e.g.* in [37, 43]. The anisotropic case $n \geq 2$ is less standard, but it will easily follow from the same arguments.

Namely, we assume that the mutation effects on phenotypes follow a normal distribution $\mathcal{N}(0, \Lambda)$, with $\Lambda = \mathrm{diag}\,(\lambda_1, \dots, \lambda_n)$ and $\lambda_i > 0$ for each $i \in [\![1, n]\!]$, and that these mutations occur with a rate $U > 0$. In other words, the dynamics of the phenotype distribution under the mutation effects only (*i.e.*, without selection) can be described by an integro-differential equation:

$$\partial_t q(t, \mathbf{x}) = U\,(J \star q - q)(t, \mathbf{x}),\ t \geq 0,\ \mathbf{x} \in \mathbb{R}^n,$$

where $\star$ is the standard convolution product in $\mathbb{R}^n$ defined by:

$$(J \star q)(t, \mathbf{x}) = \int_{\mathbb{R}^n} q(t, x_1 - y_1, \dots, x_n - y_n)\, J(y_1, \dots, y_n)\, dy_1 \dots dy_n,$$

and J the (Gaussian) probability density function associated with the normal distribution $\mathcal{N}(0, \Lambda)$.

Formally, by writing a Taylor expansion of $q(t, \mathbf{x} - \mathbf{y})$ at $\mathbf{x}$:

$$q(t, \mathbf{x} - \mathbf{y}) = \sum_{k_1, \dots, k_n = 0}^{\infty} (-1)^{k_1 + \dots + k_n} \frac{y_1^{k_1} \cdots y_n^{k_n}}{k_1! \cdots k_n!} \frac{\partial^{k_1 + \dots + k_n} q}{\partial x_1^{k_1} \cdots \partial x_n^{k_n}}(t, \mathbf{x}),$$

and by defining the central moments of the normal distribution:

$$\omega_{k_1, \dots, k_n} = \int_{\mathbb{R}^n} y_1^{k_1} \cdots y_n^{k_n}\, J(y_1, \dots, y_n)\, dy_1 \dots dy_n,$$

we then get that:

$$(J \star q)(t, \mathbf{x}) = \sum_{k_1, \dots, k_n = 0}^{\infty} (-1)^{k_1 + \dots + k_n} \frac{\omega_{k_1, \dots, k_n}}{k_1! \cdots k_n!} \frac{\partial^{k_1 + \dots + k_n} q}{\partial x_1^{k_1} \cdots \partial x_n^{k_n}}(t, \mathbf{x}).$$

Since $\omega_{k_1, \dots, k_n} = 0$ if at least one of the k_i's is odd, and since:

$$\omega_{k_1, \dots, k_n} = \Big(\prod_{i=1}^{n} \frac{k_i!}{2^{k_i/2} (k_i/2)!} \Big) \times \Big(\prod_{i=1}^{n} \lambda_i^{k_i/2} \Big),$$

otherwise, one infers in particular that the second-order moments with even indexes are such that $\omega_{0, \dots, 0, k_i = 2, 0, \dots, 0} = \lambda_i$. Assuming that $\max_{1 \leq i \leq n} \lambda_i << 1$, we may formally neglect the moments of order $k_1 + \dots + k_n \geq 4$, leading to:

$$(J \star q)(t, \mathbf{x}) \approx q(t, \mathbf{x}) + \sum_{i=1}^{n} \frac{\lambda_i}{2} \partial_{ii}(t, \mathbf{x}).$$

Finally, setting $\mu_i^2 = U\,\lambda_i$, we obtain:

$$U\,(J \star q - q)(t, \mathbf{x}) \approx \sum_{i=1}^{n} \frac{\mu_i^2}{2} \partial_{ii} q(t, \mathbf{x}).$$

Acknowledgments

This work was supported by the French Agence Nationale de la Recherche (ANR-13-ADAP-0016 “Silentadapt” to G.M., ANR-13-ADAP-0006 “MeCC” to L.R., ANR-14-CE25-0013 “NON-LOCAL” to F.H. and L.R. and ANR-18-CE45-0019 “RESISTE” to F.H., F.L., G.M. and L.R.). This work was fostered by stimulating discussions with Thomas Lenormand. The authors also thank the reviewers for valuable comments and suggestions.

References

[1] M. Alfaro and R. Carles, Explicit solutions for replicator-mutator equations: Extinction versus acceleration, *SIAM Journal on Applied Mathematics* **74** (2014) 1919–1934.

[2] M. Alfaro and R. Carles, Replicator-mutator equations with quadratic fitness, *Proceedings of the American Mathematical Society* **145** (2017) 5315–5327.

[3] M. Alfaro and M. Veruete, Evolutionary branching via replicator-mutator equations, *Journal of Dynamics and Differential Equations* (2018) 1–24.

[4] D. G. Aronson and P. Besala, Parabolic equations with unbounded coefficients, *Journal of Differential Equations* **3** (1967) 1–14.

[5] E. Baake, A. G. Casanova, S. Probst and A. Wakolbinger, Modelling and simulating Lenski’s long-term evolution experiment, *Theoretical population biology* **127** (2019) 58–74.

[6] G. Barles, S. Mirrahimi and B. Perthame, Concentration in Lotka-Volterra parabolic or integral equations: a general convergence result, *Methods and Applications of Analysis* **16** (2009) 321–340.

[7] A. G. Casanova, N. Kurt, A. Wakolbinger and L. Yuan, An individual-based model for the Lenski experiment, and the deceleration of the relative fitness, *Stochastic Processes and their Applications* **126** (2016) 2211–2252.

[8] N. Champagnat, R. Ferrière and S. Méléard, Unifying evolutionary dynamics: from individual stochastic processes to macroscopic models, *Theoretical population biology* **69** (2006) 297–321.

[9] V. A. de Crécy-Lagard, J. Bellalou, R. Mutzel and P. Marlière, Long term adaptation of a microbial population to a permanent metabolic constraint: overcoming thymineless death by experimental evolution of *Escherichia coli*, *BMC biotechnology* **1** (2001) 10.

[10] O. Diekmann, P.-E. Jabin, S. Mischler and B. Perthame, The dynamics of adaptation: an illuminating example and a Hamilton–Jacobi approach, *Theoretical population biology* **67** (2005) 257–271.

[11] S. Elena and R. Sanjuán, Climb every mountain?, *Science* **302** (2003) 2074–2075.

[12] S. S. Fong and B. Ø. Palsson, Metabolic gene-deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes, *Nature genetics* **36** (2004) 1056.

[13] C. Fraïsse and J. J. Welch, The distribution of epistasis on simple fitness landscapes, *Biology letters* **15** (2019) 20180881.

[14] A. Friedman, *Partial Differential Equations of Parabolic Type* (Prentice-Hall, Englewood Cliffs, NJ, 1964).

[15] S. Gandon and S. Mirrahimi, A Hamilton–Jacobi method to describe the evolutionary equilibria in heterogeneous environments and with non-vanishing effects of mutations, *Comptes Rendus Mathematique* **355** (2017) 155–160.

[16] P. J. Gerrish and P. D. Sniegowski, Real time forecasting of near-future evolution, *Journal of the Royal Society Interface* **9** (2012) 2268–2278.

[17] M.-E. Gil, F. Hamel, G. Martin and L. Roques, Mathematical properties of a class of integro-differential models from population genetics, *SIAM J Appl Math* **77** (2017) 1536–1561.

[18] M.-E. Gil, F. Hamel, G. Martin and L. Roques, Dynamics of fitness distributions in the presence of a phenotypic optimum: an integro-differential approach, *Nonlinearity* (2019) in press.

[19] B. H. Good and M. M. Desai, The impact of macroscopic epistasis on long-term evolutionary dynamics, *Genetics* **85** (2015) 177–190.

[20] P. D. Hislop and I. M. Sigal, *Introduction to spectral theory: With applications to Schrödinger operators*, volume 113 (Springer Science & Business Media, 2012).

[21] S. Kryazhimskiy, G. Tkačik and J. B. Plotkin, The dynamics of adaptation on correlated fitness landscapes, *Proceedings of the National Academy of Sciences* **106** (2009) 18638–18643.

[22] R. A. LaCroix, t. E. Sandberg, E. J. O'Brien, J. Utrilla, A. Ebrahim, G. I. Guzman, R. Szubin, B. O. Palsson and A. M. Feist, Use of adaptive laboratory evolution to discover key mutations enabling rapid growth of *Escherichia coli* K-12 MG1655 on glucose minimal medium, *Applied and Environmental Microbiology* **81** (2015) 17–30.

[23] R. E. Lenski, M. R. Rose, S. C. Simpson and S. C. Tadler, Experimental evolution in *Escherichia coli.* I. Adaptation and divergence during 2,000 generations, *The American Naturalist* **138** (1991) 1315–1341.

[24] R. E. Lenski and M. Travisano, Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations, *Proceedings of the National Academy of Sciences* **91** (1994) 6808–6814.

[25] A. Lorz, S. Mirrahimi and B. Perthame, Dirac mass dynamics in multidimensional nonlocal parabolic equations, *Communications in Partial Differential Equations* **36** (2011) 1071–1098.

[26] A. Lunardi, Schauder theorems for linear elliptic and parabolic problems with unbounded coefficients in $\mathbb{R}^n$, *Studia Mathematica* **128** (1998) 171–198.

[27] G. Martin, Fisher's geometrical model emerges as a property of complex integrated phenotypic networks, *Genetics* **197** (2014) 237–255.

[28] G. Martin, S. F. Elena and T. Lenormand, Distributions of epistasis in microbes fit predictions from a fitness landscape model, *Nature Genetics* **39** (2007) 555.

[29] G. Martin and T. Lenormand, A general multivariate extension of Fisher's geometrical model and the distribution of mutation fitness effects across species, *Evolution* **60** (2006) 893–907.

[30] G. Martin and T. Lenormand, The fitness effect of mutations across environments: a survey in light of fitness landscape models, *Evolution* **60** (2006) 2413–2427.

[31] G. Martin and T. Lenormand, The fitness effect of mutations across environments: Fisher's geometrical model with multiple optima, *Evolution* **69** (2015) 1433–1447.

[32] G. Martin and L. Roques, The non-stationary dynamics of fitness distributions: Asexual model with epistasis and standing variation, *Genetics* **204** (2016) 1541–1558.

[33] I. S. Novella, E. A. Duarte, S. F. Elena, A. Moya, E. Domingo and J. J. Holland, Exponential increases of RNA virus fitness during large population transmissions, *Proceedings of the National Academy of Sciences* **92** (1995) 5841–5844.

[34] L. Perfeito, A. Sousa, T. Bataillon and I. Gordo, Rates of fitness decline and rebound suggest pervasive epistasis, *Evolution* **68** (2014) 150–162.

[35] B. Perthame and G. Barles, Dirac concentrations in Lotka-Volterra parabolic PDEs, *Indiana University Mathematics Journal* (2008) 3275–3301.

[36] M. H. Protter and H. F. Weinberger, *Maximum Principles in Differential Equations* (Prentice-Hall, Englewood Cliffs, NJ, 1967).

[37] L. Roques, *Modèles de réaction-diffusion pour l'écologie spatiale* (Editions Quae, 2013).

[38] F. Rosenzweig and G. Sherlock, Experimental evolution: prospects and challenges, *Genomics* **104** (2014) v.

[39] S. Schoustra, S. Hwang, J. Krug and J. A. G. de Visser, Diminishing-returns epistasis among random beneficial mutations in a multicellular fungus, *Proceedings of the Royal Society B: Biological Sciences* **283** (2016) 20161376.

[40] P. D. Sniegowski and P. J. Gerrish, Beneficial mutations and the dynamics of adaptation in asexual populations, *Philosophical Transactions of the Royal Society B: Biological Sciences* **365** (2010) 1255–1263.

[41] O. Tenaillon, The utility of Fisher's geometric model in evolutionary genetics, *Annual Review of Ecology, Evolution, and Systematics* **45** (2014) 179–201.

[42] L. S. Tsimring, H. Levine and D. A. Kessler, RNA virus evolution via a fitness-space model, *Physical review letters* **76** (1996) 4440–4443.

[43] P. Turchin, *Quantitative Analysis of Movement: Measuring and Modeling Population Redistribution in Animals and Plants* (Sinauer, Sunderland, MA, 1998).

[44] D. Waxman and J. R. Peck, Pleiotropy and the preservation of perfection, *Science* **279** (1998) 1210–1213.

[45] M. J. Wiser, N. Ribeck and R. E. Lenski, Long-term dynamics of adaptation in asexual populations, *Science* (2013) 1243357.

[46] A. L. Yakymiv, A generalization of the Curtiss theorem for moment generating functions, *Math. Notes* **90** (2011) 920–924.